OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

# Does Portfolio Optimization Beat Splitting Evenly?

Adrian Erlikhman and Ryan Erlikhman
erlikhman.adrian@gmail.com, rerlikhman@gmail.com

## 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 earns more return per unit of risk. 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. This gap is both statistically significant and economically large: at the 16% volatility of the 1/*N* portfolio it is worth about 3 percentage points of return per year, and an approximate adjustment for trading costs of 50 basis points per dollar traded roughly doubles it. No strategy outperforms the 1/*N* portfolio; the best performers ignore the estimated expected returns for the assets, and the only statistically detectable gap between them and 1/*N* (0.04 in Sharpe ratio) is too small to matter economically.

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, although trading costs could erase the smallest of these advantages. 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.

## 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, $\mathbf{\Sigma}^{-1}\boldsymbol{\mu}$. 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. Throughout, we distinguish statistical significance, which asks whether a difference between two strategies is larger than the noise in our simulations, from economic significance, which asks whether it is large enough to matter to an investor, and we examine how transaction costs would change each comparison.

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 $\mathbf{w} \in \mathbb{R}^N$. The constraint that the weights sum to one means that $\mathbf{w}$ must satisfy $\mathbf{w}^\top \mathbf{1} = 1$, where $\mathbf{1}$ denotes a vector of $N$ ones. Suppose further that the assets have expected excess returns $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, so that the portfolio has expected excess return $\mathbf{w}^\top \boldsymbol{\mu}$ and variance $\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}$. Then the mean–variance investor with risk aversion $\gamma$ aims to solve

$$\max_{\mathbf{w}:\, \mathbf{w}^\top \mathbf{1}=1} \quad \mathbf{w}^\top \boldsymbol{\mu} - \frac{\gamma}{2} \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \tag{1}$$

Two portfolios are central to this problem. The first is the well-known tangency (or "maximum-Sharpe") portfolio

$$\mathbf{w}_{MV} = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}} \tag{2}$$

which has the highest Sharpe ratio, $\mathbf{w}^\top \boldsymbol{\mu} / (\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w})^{1/2}$, of any fully invested portfolio whenever $\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} > 0$. The second arises 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 the solution of Eq. (1) as risk aversion $\gamma$ grows without bound. It is the global minimum-variance (GMV) portfolio

$$\mathbf{w}_{GMV} = \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}} \tag{3}$$

For any finite $\gamma$, the solution of Eq. (1) mixes these two portfolios,

$$\mathbf{w}^*(\gamma) = \mathbf{w}_{GMV} + \frac{b}{\gamma}\left(\mathbf{w}_{MV} - \mathbf{w}_{GMV}\right), \qquad b = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \tag{4}$$

so the tangency portfolio is the solution of Eq. (1) for the particular risk aversion $\gamma = b$ (when $b > 0$). The unconstrained mean–variance rules we evaluate use the tangency weights of Eq. (2) directly, so their weights do not depend on a choice of $\gamma$.

The comparison between these two portfolios is the main experimental lever of this paper. The GMV portfolio requires only the covariance matrix $\boldsymbol{\Sigma}$, whereas the tangency portfolio additionally requires the vector $\boldsymbol{\mu}$ 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 $\boldsymbol{\mu}$: whenever the GMV portfolio outperforms the tangency portfolio in practice, that gap is attributable to estimation error in $\boldsymbol{\mu}$.

### 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 $Y$ years of data is $\sigma/\sqrt{Y}$, so estimating $\mu$ with a standard

error of just 2% per year requires $Y = (20/2)^2 = 100$ years of returns (we write $Y$ for a length of history in years, to keep it distinct from the estimation window $M$, which is measured in months). 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 $\mathbf{\Sigma}^{-1}\boldsymbol{\mu}$ 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 ($\mathbf{w} \geq \mathbf{0}$, element by element) 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

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{B}\,\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \qquad \mathbf{f}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_f\right), \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \tag{5}$$

where $\mathcal{N}$ denotes the normal distribution, the factor loadings $\mathbf{B}$ are centered at one, and the idiosyncratic variances on the diagonal of $\mathbf{D}$ 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 $\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^\top + \mathbf{D}$ and mean $\boldsymbol{\mu}$, 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 ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$), 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.

## Strategies

Table 1 summarizes the seven rules that we compare. For each rule, we first form an estimate of the average return within the estimation window, denoted $\hat{\boldsymbol{\mu}}$, and an estimate of the covariance matrix of asset returns within that window, denoted $\hat{\boldsymbol{\Sigma}}$. We then plug these values into one of the weight formulas given above (Eqs. (2) and (3)), optionally substituting shrunk estimates of the mean or covariance matrix.

**Table 1:** 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 | $w_i = 1/N$ |

| Strategy | Inputs used | Weight rule |
|---|---|---|
| MV (sample) | $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$ | Eq. (2) |
| MV (long-only) | $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}$ | Eq. (2) subject to $\mathbf{w} \geq \mathbf{0}$ |
| MV (LW cov) | $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{LW}}$ | Eq. (2) with shrunk $\hat{\boldsymbol{\Sigma}}$ |
| MV (LW cov + BS mean) | $\hat{\boldsymbol{\mu}}_{\mathrm{BS}}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{LW}}$ | Eq. (2), both shrunk |
| Min-Var (sample) | $\hat{\boldsymbol{\Sigma}}$ | Eq. (3) |
| Min-Var (LW cov) | $\hat{\boldsymbol{\Sigma}}_{\mathrm{LW}}$ | Eq. (3) with shrunk $\hat{\boldsymbol{\Sigma}}$ |

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 A). 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 $\mathbf{w}_t$, and record the portfolio return $\mathbf{w}_t^{\top}\mathbf{r}_t$ in the next month. No information from month $t$ or any later month is used to form $\mathbf{w}_t$, 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,

$$SR = \sqrt{12}\,\frac{\bar{r}}{s_r} \tag{6}$$

where $\bar{r}$ 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, $\sigma = \sqrt{12}\,s_r$; the certainty-equivalent return

$$CEQ = 12\left(\bar{r} - \frac{\gamma}{2}\,s_r^2\right) \tag{7}$$

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. Formally, following DeMiguel et al. (2009), turnover is

$$\tau = \frac{1}{T}\sum_{t}\sum_{j=1}^{N}\left|w_{j,t+1} - w_{j,t}^{+}\right| \tag{8}$$

where $w_{j,t}^{+}$ is the weight of asset $j$ just before rebalancing at the end of month $t$—after that month's returns have moved it away from its target—and the average is taken over the $T$ monthly rebalancing dates of the out-of-sample period. Because it counts both purchases and sales, a turnover of 0.05 means that trades worth 5% of the portfolio's value are needed each month.

## 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.

## Statistical Versus Economic Significance

Each comparison is evaluated in two different ways. The first is *statistical significance*: whether the difference between strategies is large relative to the noise in the simulations. In Experiment A we evaluate the difference between the out-of-sample Sharpe ratio of $1/N$ and the out-of-sample Sharpe ratio of each other strategy in each simulation. We test whether the paired differences are centered on zero using a two-sided Wilcoxon signed-rank test. We also report a paired $t$-test for the main comparison. These paired tests are quite sensitive, since all strategies experience identical return paths and universes. However, a $p$-value from a simulation study provides information about the precision of a Monte Carlo estimate rather than the magnitude of an effect. The number of simulations is set by the analyst. Increasing the number of simulations will drive the $p$-value of any nonzero effect, however small, towards zero, without increasing the size of the effect.

The second evaluation is *economic significance*: is the difference large enough to be important to an investor? We evaluate this in three different ways. First we convert differences in Sharpe ratios into differences in expected returns at a given level of risk. Levering a portfolio up or down (by borrowing or lending at the risk-free rate) does not alter the Sharpe ratio of the portfolio. Thus two strategies with the same level of volatility $\sigma$ have a difference in expected excess return of $\Delta SR \times \sigma$. We use the volatility of the $1/N$ portfolio in Table 2 ($\sigma = 16\%$). At this level of volatility $1/N$ is expected to outperform the risk-free rate by $0.70 \times 16\% \approx 11.2\%$ per year. Thus a difference in Sharpe ratios of 0.10 corresponds to a difference in expected returns of 1.6 percentage points per year. Second we consider Pr($1/N$ wins). This is the fraction of simulations in which $1/N$ comes out ahead, which is what an investor who lives through only one market history can expect. Finally we compare the differences to the uncertainty in a single investor's experience. According to Lo (2002), for normally distributed and independent returns (as in our simulations), the standard error in a Sharpe ratio based on $T$ observations is approximately $\sqrt{((1 + SR^2/2)/T)}$, where SR is measured on a periodic basis. If the annualized Sharpe ratio is estimated using monthly returns over $Y$ years then

$$SE\left(\hat{SR}\right) \approx \sqrt{\frac{1+SR^2/24}{Y}} \quad (9)$$

From our 15-year study period and at SR = 0.70, Eq. (9) implies ≈ 0.26. This would mean that an individual investor would not be able to reliably identify two investment strategies with a small difference (a few hundredths) in their Sharpe ratio, no matter how small the *p*-value in our simulations. Dividing this 0.26 by √250 yields ≈ 0.02, which serves as a lower bound of the Monte Carlo standard error associated with the mean Sharpe ratio for a strategy in Experiment A (this is a lower bound since the actual Sharpe ratios also differ across universes). Strategies whose simulation results exhibit more fluctuation across simulations, such as sample Markowitz in Figure 2, will have larger standard errors. Figure 1 displays the precision of each mean using 95% intervals. Another economic metric is the certainty-equivalent return in Eq. (7). At $\gamma = 1$ it rewards higher volatility as well as a higher Sharpe ratio, so we use it mainly to flag the extreme risks of the unconstrained strategies. We define a difference of a few hundredths in Sharpe ratio as economically small (this represents less than one percentage point of annual return at 16% volatility), a difference between about 0.05 and 0.10 as economically moderate, and a difference of 0.10 or greater as economically meaningful. Transaction costs are discussed next and are part of the economic side of the comparison. Transaction costs do affect the magnitude of the above differences and can reverse an investor's preference even when a difference exists prior to transaction costs and is statistically significant.

## Transaction Costs

Our simulations assume that trading is free. To gauge how costs would change the results, suppose that every dollar traded costs a fixed proportion $c$ of its value, covering commissions, bid–ask spreads, and price impact. The return of a strategy in month $t + 1$ net of costs is then approximately

$$r_{t+1}^{net} = \mathbf{w}_{t+1}^{\top}\mathbf{r}_{t+1} - c\sum_{j=1}^{N}\left|w_{j,t+1} - w_{j,t}^{+}\right| \quad (10)$$

where $\mathbf{w}_{t+1}^{\top}\mathbf{r}_{t+1}$ is the portfolio's return before costs. Costs therefore lower the mean monthly return by about $c\tau$, and the annual mean by $12c\tau$, while leaving the volatility almost unchanged, so the Sharpe ratio net of costs is approximately

$$SR_{net} \approx SR - \frac{12\,c\,\tau}{\sigma} \quad (11)$$

where $\sigma$ is the annualized volatility. What matters is the ratio $\tau/\sigma$, turnover per unit of risk: levering a portfolio up multiplies its turnover and its volatility by roughly the same factor, so a rule is penalized not for its leverage as such but for how much its composition changes from month to month. A rule $k$ that beats $1/N$ before costs by $\Delta SR > 0$, and trades more per unit of risk than $1/N$ ($\tau_k/\sigma_k > \tau_{1/N}/\sigma_{1/N}$), keeps its advantage only while $c$ stays below the break-even cost

$$c^{*} = \frac{\Delta SR}{12\left(\tau_k/\sigma_k - \tau_{1/N}/\sigma_{1/N}\right)} \quad (12)$$

We evaluate Eq. (11) at $c$ = 10, 25, and 50 basis points (0.10%, 0.25%, and 0.50% of each dollar traded); because the cost term is proportional to $c$, readers can rescale the results to any other cost level. Because we apply Eq. (11) to the averages reported in Table 2 rather than deducting costs month by month inside each simulation, the results are approximations, but they are transparent and use only quantities that we report.

## RESULTS

### The Headline: Naive Beats Optimal

Table 2 and Figure 1 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 statistically highly significant (Wilcoxon $p \approx 3 \times 10^{-14}$, paired $t$-test $p \approx 4 \times 10^{-4}$; the $t$-test is less decisive because the two extreme Markowitz outcomes in Figure 2 account for almost all of the variance of the paired differences), and 1/$N$ wins in 74% of individual simulations. It is also economically large. At the 16% volatility of 1/$N$, a Sharpe ratio gap of 0.19 is worth about 3.0 percentage points of expected return per year—roughly $300 a year on a $10,000 portfolio—so the Markowitz portfolio earns about 27% less than 1/$N$ for the same risk. Even this gap is smaller than the 0.26 standard error of a single 15-year track record (Eq. (9)), and the Markowitz outcomes are more dispersed still (Figure 2), so the Markowitz rule comes out ahead in about one simulation in four: the average investor is clearly worse off, but any one investor's history would be too noisy to show it reliably.

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; they also tie economically, since they trail 1/$N$ by at most 0.03 (about 0.5 percentage points per year at 16% volatility) and 1/$N$ wins in only half of the simulations. By CEQ, MV (long-only) is ahead of 1/$N$ (0.112 against 0.098), but only because an investor with $\gamma$ = 1 welcomes its higher volatility (0.19 against 0.16); at equal volatility the two portfolios would have the same expected return. The three rules based on unconstrained estimates of expected returns all have Sharpe ratios between 0.50 and 0.59, gaps of 0.11 to 0.20 that are both statistically significant ($p \leq 6 \times 10^{-10}$) and economically meaningful (1.8 to 3.2 percentage points per year). 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. A gap of 0.04 is worth about 0.6 percentage points per year at 16% volatility, is about one-sixth of the uncertainty in a single 15-year track record, and leaves Min-Var (LW cov) ahead of 1/$N$ in 40% of the simulations: it is statistically real but economically small. 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.

**Table 2:** 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; volatility and CEQ are annual figures in decimal form (0.16 = 16% per year). 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 signed-rank 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. Min-Var (LW cov) differs from 1/*N* statistically but trails it by only 0.04, an economically small gap. 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 | $3 \times 10^{-14}$ |
| 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 | $2 \times 10^{-19}$ |
| MV (LW cov + BS mean) | 0.59 | 1.46 | −14.9 | 4.87 | 0.68 | $6 \times 10^{-10}$ |
| 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 | $2 \times 10^{-4}$ |

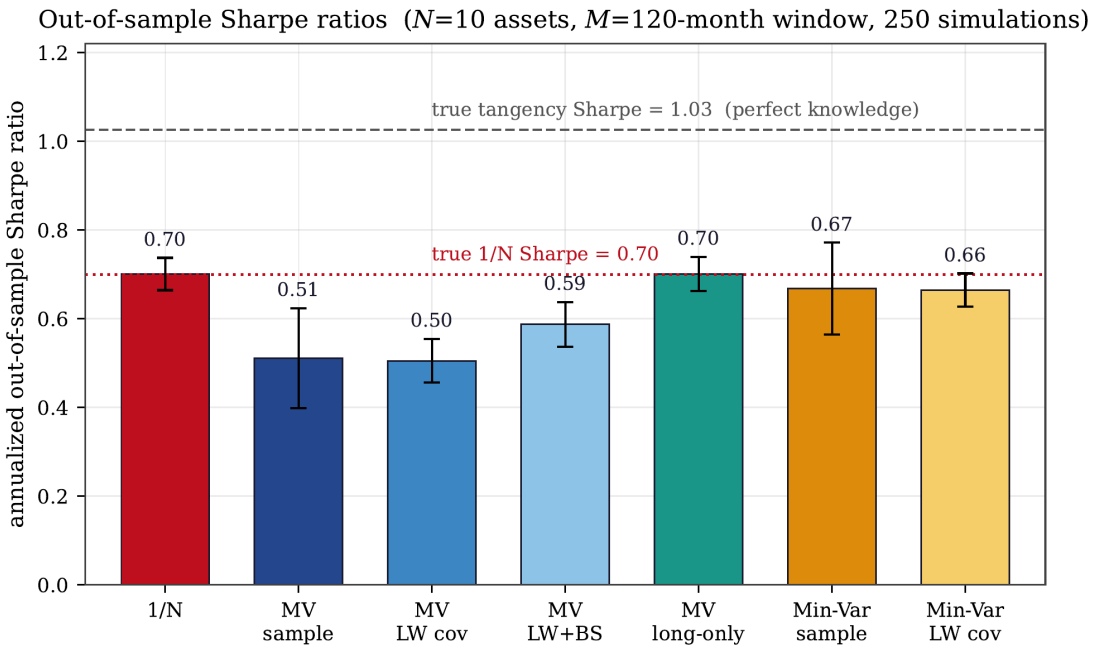


**Figure 1:** Out-of-sample Sharpe ratios by strategy (error bars are 95% confidence intervals across simulations; they show how precisely the simulations estimate each mean, not the much wider range of outcomes a single investor could experience, which Figure 2 illustrates). The dashed line is the Sharpe 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.

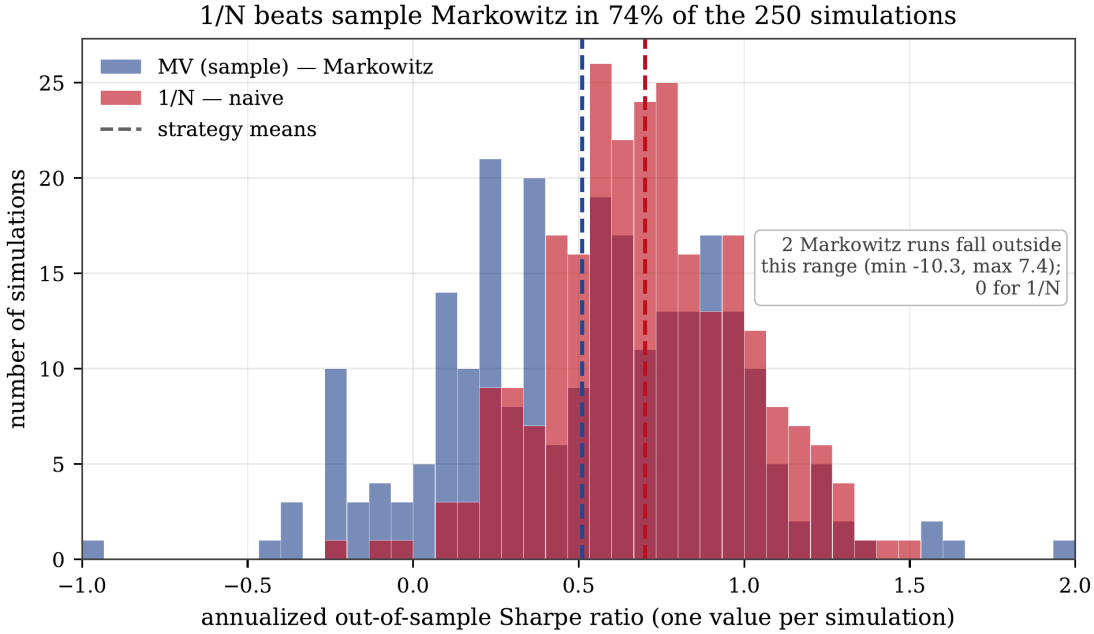


**Figure 2:** Distribution of the out-of-sample Sharpe ratio across the 250 simulations for $1/N$ (red) and the sample Markowitz rule (blue). The Markowitz distribution shows both a leftward shift and far greater dispersion, the

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 2 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 trading costs would absorb much of what the strategy earns: at 50 basis points per dollar traded they would come to about 81% of the portfolio's value each year and would roughly double its Sharpe ratio shortfall relative to 1/*N* (see "How Transaction Costs Would Change the Results" below). Figure 3 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.

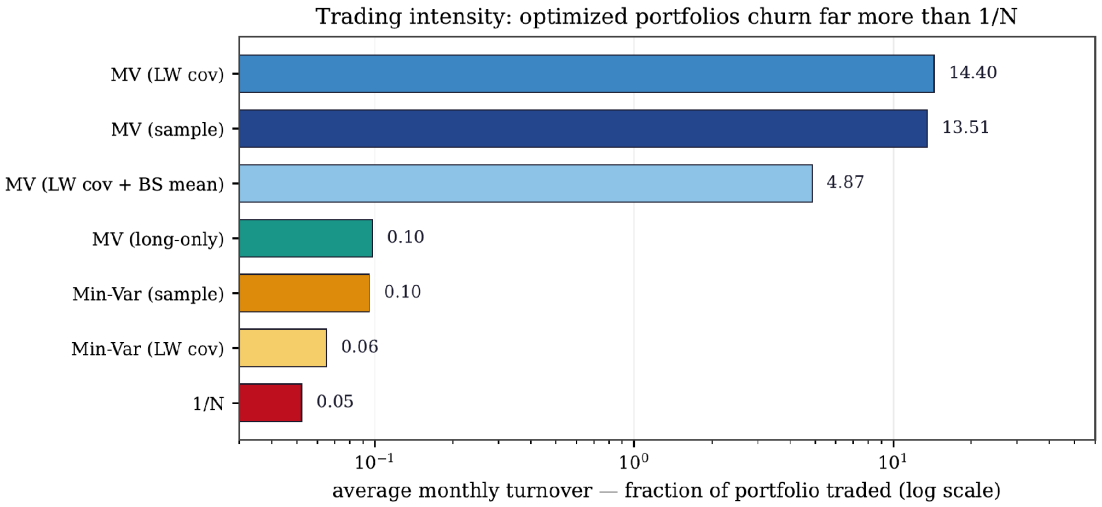


**Figure 3:** Average monthly turnover (log scale). The optimized portfolios that use estimated means trade one to two orders of magnitude more than 1/*N* or the minimum-variance rules, because their weights chase estimation noise from month to month.

## The Mechanism: It Is the Means

Experiment C settles the diagnosis. Figure 4 feeds the tangency formula (Eq. (2)) 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 2; the difference is Monte Carlo noise of the expected size, about one Monte Carlo standard error, 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 plays only a secondary role. In economic terms, estimating the covariance matrix costs 0.08 in Sharpe ratio (about 1.3 percentage points of return per year at 16% volatility, a moderate amount), whereas estimating the expected returns costs 0.38 (about 6.1 percentage points per year), nearly five times as much. We did not run formal tests in Experiment C, and the error bars in Figure 4 show that the individual means are estimated with considerable uncertainty; the 0.38 cost of estimating expected returns is large relative to that uncertainty, whereas the 0.08 cost of estimating the covariance matrix is not.

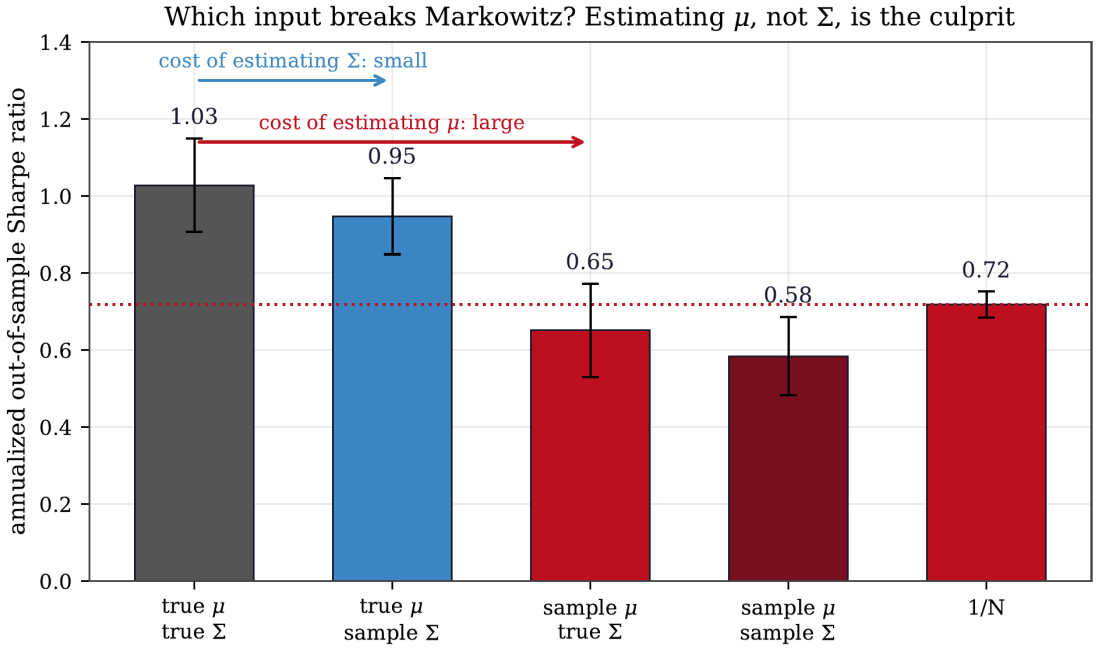


**Figure 4:** Oracle decomposition (Experiment C). Each bar feeds the Markowitz formula a different combination of true and estimated inputs. Estimating the covariance matrix (second bar) costs comparatively little; estimating expected returns (third bar) is what breaks the optimizer, dragging it below 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 2 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, an economically negligible change of less than 0.2 percentage points of return per year at 16% volatility. 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 $50 \times 50$ 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 3 and Figure 5). Improvements of this size, about 0.3 in Sharpe ratio or roughly 5 percentage points of return per year at 16% volatility, are economically large by any standard. 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 2, 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, a gain worth about 1.4 percentage points per year at 16% volatility). 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 trails by 0.11, a gap that is both statistically significant, $p = 6 \times 10^{-10}$, and economically meaningful), 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 $1/N$ ($p = 0.68$) and economically indistinguishable from it: the two Sharpe ratios agree to two decimal places, and $1/N$ wins in half of the simulations. 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).

**Table 3:** Sensitivity grid (Experiment B): mean out-of-sample Sharpe ratios across 200 simulations per cell. "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. No formal tests were run for this grid. As a guide to Monte Carlo noise, the row $N = 10$, $M = 120$ can be compared with the independent simulations of Table 2 (0.68 against 0.70 for $1/N$, and 0.52 against 0.51 for sample Markowitz), so differences of a few hundredths should not be over-interpreted.

| *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 |
| 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 |

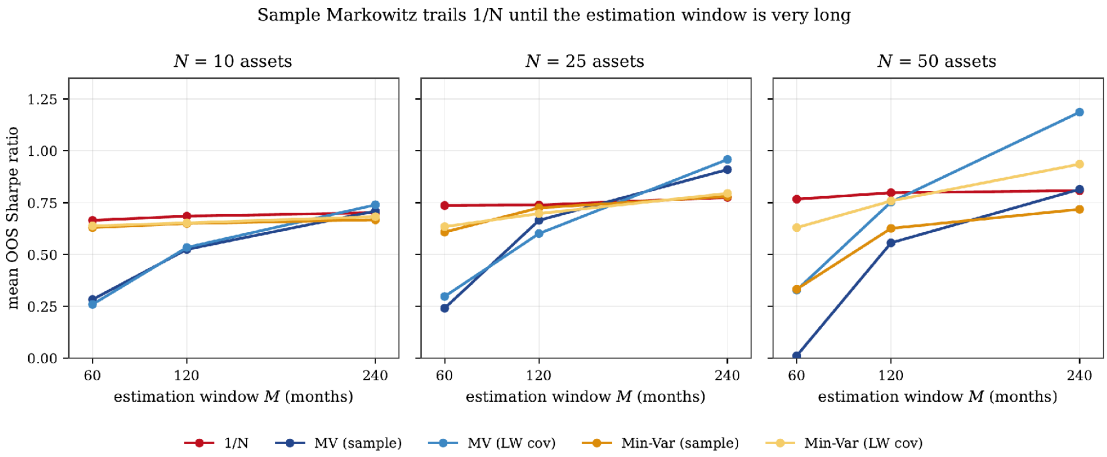


**Figure 5:** Out-of-sample Sharpe ratio versus estimation-window length $M$, for $N = 10$, 25, 50 assets (all panels share a common vertical scale). Sample Markowitz (dark blue) starts far below $1/N$ (red) and climbs as more data become available, reaching $1/N$ only around a 20-year window. The collapse 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 5 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 $N = 50$. These margins differ in economic importance. The edge of 0.04 at $N = 10$ is small—about 0.6 percentage points per year at 16% volatility, untested statistically, and small enough that modest trading costs could erase it—whereas the edges of 0.19 and 0.38 at $N = 25$ and $N = 50$ are economically large, although they arise in universes whose attainable Sharpe ratios (1.94 and 3.73) are unrealistically high, which inflates the prize for optimizing. 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.

## How Transaction Costs Would Change the Results

Table 4 uses Eq. (11) with the averages in Table 2. Three observations follow.

**Table 4:** Approximate Sharpe ratios net of proportional transaction costs (Experiment A) from Eq. (11) and the averages in Table 2. $c$ is the cost per dollar traded in basis points (1 basis point = 0.01%); $\tau/\sigma$ is average monthly turnover divided by annualized volatility. The final column is the annual cost at $c$ = 50 basis points, $12c\tau$, as a percentage of the portfolio value. These values are particularly rough for the unconstrained Markowitz rules because their averages are dominated by a few extreme simulations.

| Strategy | $\tau/\sigma$ | $c = 0$ | 10 bp | 25 bp | 50 bp | Annual cost at 50 bp |
|---|---|---|---|---|---|---|
| $1/N$ | 0.31 | 0.70 | 0.70 | 0.69 | 0.68 | 0.3% |
| MV (sample) | 3.51 | 0.51 | 0.47 | 0.40 | 0.30 | 81% |
| MV (long-only) | 0.53 | 0.70 | 0.69 | 0.68 | 0.67 | 0.6% |
| MV (LW cov) | 4.83 | 0.50 | 0.44 | 0.36 | 0.21 | 86% |
| MV (LW cov + BS mean) | 3.34 | 0.59 | 0.55 | 0.49 | 0.39 | 29% |
| Min-Var (sample) | 0.67 | 0.67 | 0.66 | 0.65 | 0.63 | 0.6% |
| Min-Var (LW cov) | 0.40 | 0.66 | 0.66 | 0.65 | 0.64 | 0.4% |

1. Transaction costs only increase the advantage of $1/N$ in Experiment A. All other strategies trade more per unit of risk than does $1/N$ (their $\tau/\sigma$ values range from 0.40 to 4.83 versus 0.31 for $1/N$), so all their differences from $1/N$ increase with $c$. For strategies that are tied or nearly tied to $1/N$, the increases are economically small: at 50 basis points MV (long-only) and both of the minimum-variance strategies trail $1/N$ by between 0.01 and 0.05.

2. For strategies using unconstrained estimates of expected returns, transaction costs are severe: at 50 basis points they multiply all differences from $1/N$ by two or three, to 0.38, 0.47, and 0.29. The sample Markowitz strategy would have an annual trading cost equal to 81% of its portfolio value. Its Sharpe ratio would fall from 0.51 to approximately 0.30, doubling its difference from $1/N$ (from 0.19 to 0.38). The Ledoit–Wolf covariance strategy does even worse (0.21) because it trades the most per unit of risk. Even the doubly shrunk strategy falls to 0.39. At just 10 basis points, these strategies suffer a drop in Sharpe ratio of 0.04 to 0.06, which is comparable to the difference between Min-Var (LW cov) and $1/N$.

3. Costs are most important in situations with the greatest potential for optimization: the long-window cases in Table 3. Experiment B does not report turnover, so Eq. (12) cannot be computed exactly for those cases, but a rough computation is useful. Under the MV (LW) rule, if the turnover per unit of risk at $M = 240$ was equal to that seen in Experiment A ($\tau/\sigma \approx 4.8$), then a cost of approximately 7 basis points would eliminate the 0.04 advantage at $N = 10$, a cost of approximately 35 basis points would eliminate the advantage at $N = 25$, and a cost of approximately 70 basis points would eliminate the advantage at $N = 50$. Longer windows should help to stabilize the weights and decrease turnover, but having more assets should work in the opposite direction. If turnover per unit of risk decreased to one quarter of the level in Experiment A, then the break-even costs would increase to approximately 37, 177, and 354 basis points. Thus the small absolute advantage at $N = 10$ is quite sensitive to costs, but the larger advantages at $N = 25$ and $N = 50$ may be less sensitive to moderate costs.

All of these cost estimates are biased in favor of the optimized strategies. Missing from Eq. (10) are costs that would hit these strategies the hardest: costs associated with borrowing the securities they short, financing and margin costs on their large long and short positions (in our implementation we assume the cash generated from short sales can be freely reinvested), and market impact that increases with the size of the transaction, which means that a constant-percentage cost will be substantially understated for strategies that are trading many times the size of their portfolio value monthly. All of these costs would also work in favor of $1/N$. Decreasing the frequency of rebalancing, only trading when a weight moves outside a certain band, or including turnover costs in the optimization would decrease costs across all strategies, but we leave these improvements to future research.

## DISCUSSION

Our results are a controlled reproduction of DeMiguel et al. (2009) and a direct measurement of its cause. Four 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 $\boldsymbol{\mu}$ is a virtue when $\boldsymbol{\mu}$ 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 $1/N$—an economically large edge, although it arises in a universe with an unrealistically high attainable Sharpe ratio and would shrink once trading

costs are paid. 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.

Fourth, *statistical and economic significance must be judged separately, and trading costs belong on the economic side*. Our simulations can detect gaps as small as 0.04 in Sharpe ratio, but such gaps are worth well under one percentage point of return per year and would be hard to detect in any single investor's track record. The conclusions that matter economically rest on much larger gaps—0.11 or more between 1/*N* and the unconstrained optimizers, and about 0.3 for the covariance-shrinkage rescue at $N = 50$ and $M = 60$—and transaction costs only widen the advantage of 1/*N* in the realistic setting of Experiment A.

### 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. Our treatment of transaction costs is approximate: it applies a proportional cost to each rule's average turnover and volatility rather than deducting costs month by month inside each simulation; for Experiment B, where turnover is not reported, it gives only a rough illustration; and it omits short-sale fees, financing costs, and price impact that grows with trade size—each of which would 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.

### Future Work

It would be of interest to perform the same study on real returns from market sectors or individual countries; to deduct transaction costs month by month within each simulation, including short-sale fees and price impact, and to test rules that rebalance less often or penalize turnover; 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. Its advantage over the rules that act on unconstrained estimates of expected returns is both statistically significant and economically large—about 3 percentage points of return per year against sample Markowitz at the same level of risk—and trading costs of 50 basis points per dollar traded would roughly double it. The smaller gaps we find, a few hundredths in Sharpe ratio, are too small to matter to an investor, even where, as for Min-Var (LW cov), our simulations detect them statistically.

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 margin that is small with ten assets but grows with the number of assets, provided its trading costs stay moderate. 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.

## 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.

Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4), 36–52.

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.

## APPENDIX A: THE LEDOIT–WOLF CONSTANT-CORRELATION ESTIMATOR

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

$$f_{ii} = s_{ii}, \qquad f_{ij} = \bar{\rho}\sqrt{s_{ii}s_{jj}} \quad (i \neq j), \qquad \bar{\rho} = \frac{2}{N(N-1)}\sum_{i<j}\frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} \tag{A1}$$

The shrinkage estimator is

$$\hat{\boldsymbol{\Sigma}}_{LW} = \delta\,\mathbf{F} + (1-\delta)\,\mathbf{S}, \qquad \delta = max\left\{0,\ min\left\{1,\ \frac{\hat{\kappa}}{M}\right\}\right\} \tag{A2}$$

where the optimal intensity depends on

$$\hat{\kappa} = \frac{\hat{\pi}-\hat{\lambda}}{\hat{\eta}}, \qquad \hat{\pi} = \sum_{i,j} Asy\hat{V}ar\left(\sqrt{M}\,s_{ij}\right), \qquad \hat{\eta} = \sum_{i,j}\left(f_{ij}-s_{ij}\right)^2 \tag{A3}$$

Here $\hat{\pi}$ measures the noise in $\mathbf{S}$, $\hat{\eta}$ measures the misfit of the target, and $\hat{\lambda}$ corrects for the covariance between the estimation errors in $\mathbf{S}$ and $\mathbf{F}$ (Ledoit and Wolf, 2004). We write $\bar{\rho}$, $\hat{\eta}$, and $\hat{\lambda}$ for the quantities that Ledoit and Wolf denote $\bar{r}$, $\hat{\gamma}$, and $\hat{\rho}$, to avoid clashes with the mean return $\bar{r}$ in Eq. (6), the risk-aversion coefficient $\gamma$, and the average correlation $\bar{\rho}$. Intuitively, $\delta$ 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 $\delta \approx 0.48$.