Modern Finance

Modern Finance Institute

Article

Simple versus optimized ETF portfolios: Does complexity pay?

Michal Mešťan [1,*]

[1] Department of Finance and Accounting, Faculty of Economics, Matej Bel University in Banská Bystrica, Tajovského 10, 975 90 Banská Bystrica, Slovakia

* Correspondence: michal.mestan@umb.sk.

Abstract: This study tests whether allocation complexity pays for long-horizon ETF investors. The full sample comprises 887 weekly observations for four equity and three bond ETFs over 2009–2025, with out-of-sample evaluation beginning in 2012 after estimation. Eleven fixed stock–bond allocations, 1/N, capped 1/SD, and constrained Markowitz optimization are compared using 1,000 paired bootstrap paths over 10–40 years. All-equity portfolios maximize median terminal wealth, while high-equity fixed allocations attain the highest median certainty equivalent. Markowitz achieves the highest median Sharpe and Sortino ratios and gains 0.64–0.69 percentage points in certainty equivalent over 1/N. Complexity therefore pays for risk-adjusted performance, but not universally.

Keywords: exchange-traded funds; asset allocation; portfolio optimization; naïve diversification; inverse-volatility weighting; transaction costs

JEL Classification: G11; G12; G17; C15

1. Introduction

Exchange-traded funds (ETFs) give investors inexpensive access to diversified equity and bond exposures through a small set of liquid, transparent instruments (Poterba & Shoven, 2002). The central portfolio decision is therefore less about selecting individual securities than about choosing an allocation rule. Simple rules are easy to understand and avoid parameter uncertainty, whereas statistical rules can exploit information about volatility, dependence, and expected returns. That information may improve portfolio efficiency, but estimation error, unstable weights, turnover, and implementation costs can erode the gain. The practical question is whether additional complexity improves the outcome an investor actually seeks.

This paper evaluates that question on a common universe of four equity and three bond ETFs. It compares eleven fixed stock–bond allocations from 100/0 to 0/100, equal weighting (1/N), inverse-volatility weighting (1/SD) capped at 30% per ETF, and constrained Markowitz optimization. The strategies are assessed in a historical out-of-sample test and on 1,000 paired moving-block bootstrap paths evaluated after 10, 20, 30, and 40 years, with annual rebalancing and explicit transaction costs.

The results reveal a clear distinction between wealth maximization and risk-adjusted efficiency. The all-equity allocation produces the highest median terminal wealth at every simulated horizon. High-equity fixed portfolios also achieve the highest median certainty equivalent: 90/10 and 100/0 are economically indistinguishable at 10 years, whereas 100/0 leads at 20, 30, and 40 years. These outcomes, however, come with substantially higher volatility and drawdowns.

Markowitz, by contrast, achieves the highest median Sharpe and Sortino ratios at every horizon and produces median paired certainty-equivalent (CEQ) gains of 0.64–0.69 percentage points relative to 1/N. The probability of a CEQ improvement over 1/N

Check for updates

Citation: Mešťan, M. (2026). Simple versus optimized ETF portfolios: Does complexity pay? *Modern Finance*, *4*(4), 1-12.

Accepting Editor: Adam Zaremba

Received: 26 July 2026
Accepted: 15 October 2026
Published: 24 October 2026

Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) license (https://creativecommons.org/licenses/by/4.0/).

increases from 67.3% at 10 years to 82.6% at 40 years. The capped 1/SD rule reduces risk but delivers lower median utility, and transaction-cost sensitivity does not alter the principal rankings. Complexity therefore pays for risk-adjusted performance relative to balanced benchmarks, but it does not dominate simple high-equity allocations for wealth or utility.

These findings contribute to the portfolio-choice literature by showing that the apparent value of complexity depends jointly on the benchmark and the performance objective. Out-of-sample optimization studies often use 1/N as the principal naïve comparator (DeMiguel et al., 2009; Kritzman et al., 2010), although its aggregate stock–bond exposure is mechanically determined by the number of ETFs in each class. Constraints can regularize estimated portfolios (Jagannathan & Ma, 2003; Behr et al., 2013), and parsimonious volatility signals can improve portfolio management (Kirby & Ostdiek, 2012; Moreira & Muir, 2017), but those results do not establish whether adaptive rules dominate nearby fixed risk allocations. By benchmarking every rule against the complete fixed grid on common return paths, and by reporting economic magnitude, improvement probability, turnover, drawdown, and multiplicity-adjusted significance, this study identifies where complexity adds value and where it does not.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature on simple and optimized allocation rules. Section 3 describes the ETF universe and data. Section 4 presents the portfolio strategies, out-of-sample design, bootstrap procedure, performance measures, and statistical tests. Section 5 reports the historical and simulated results. Section 6 discusses the findings, their relation to the literature, and the study's limitations. Section 7 concludes.

2. Literature Review

2.1. Simple allocation and naïve diversification

Fixed stock–bond portfolios express the investor's main strategic risk choice without requiring forecasts. Annual rebalancing gives these rules a limited dynamic component by restoring target exposures, but its benefit depends on the allocation, trading rule, and cost environment (Dichtl et al., 2016; Donohue & Yip, 2003). A fixed grid is therefore a useful benchmark because it separates the value of portfolio construction from the value of choosing more equity risk.

Equal weighting is a stronger benchmark than its simplicity suggests. It avoids estimation of means, variances, and correlations, and may be rational under high model ambiguity (Pflug et al., 2012). DeMiguel et al. (2009) find that optimized models often fail to dominate 1/N out of sample once estimation error and turnover are considered. However, 1/N alone is an incomplete comparator for stock–bond allocation: its aggregate equity exposure is determined mechanically by the number of ETFs in each class, so an optimizer's apparent gain may partly reflect a different risk budget rather than superior estimation. Other evidence is more favorable when naïve and optimized information are combined or when the comparison is carefully controlled (Kritzman et al., 2010; Tu & Zhou, 2011).

2.2. Optimization, constraints, and volatility-based rules

Mean–variance theory selects weights jointly because portfolio risk depends on covariances as well as standalone volatility (Markowitz, 1952). Its empirical weakness is sensitivity to estimated inputs, especially expected returns (Michaud, 1989). Covariance shrinkage and economically meaningful constraints can regularize the solution (Ledoit & Wolf, 2003). In particular, long-only and weight constraints can improve out-of-sample performance by limiting extreme responses to sampling error (Behr et al., 2013; Jagannathan & Ma, 2003).

Inverse-volatility weighting occupies an intermediate position: it uses risk estimates but neither expected-return forecasts nor matrix inversion. It is the diagonal special case

of the broader equal-risk-contribution family, which equalizes marginal risk contributions and therefore also requires covariance information (Maillard et al., 2010). Related research reports gains from parsimonious volatility signals (Kirby & Ostdiek, 2012; Moreira & Muir, 2017), while broader out-of-sample evidence is less uniform (Cederburg et al., 2020). These findings are informative but not directly interchangeable: volatility timing scales exposure through time, whereas inverse-volatility weighting reallocates exposure across assets. Because inverse volatility can concentrate strongly in low-risk instruments, an asset cap is economically relevant; it preserves the risk-based ranking while preventing a single Treasury-bill ETF from determining the entire portfolio.

2.3. Horizon, implementation costs, and research expectation

Long horizons magnify small annual differences through compounding, but they do not eliminate sequence risk or severe equity losses (Anarkulova et al., 2022). Historical ETF records are too short for direct 30- or 40-year comparisons, so multivariate block resampling provides a transparent compromise: it preserves cross-asset dependence and short-run serial structure without imposing a fully parametric return model (Künsch, 1989).

Recent studies in Modern Finance reinforce the importance of design-specific evaluation: Page et al. (2026) show that portfolio conclusions from machine-learning signals depend on the weighting scheme, while Kunjal and Peerbhai (2026) document state-dependent ETF trading responses. Yet the strands reviewed above remain segmented. Rebalancing studies primarily compare fixed rules; tests of optimization often emphasize 1/N as the principal naïve benchmark; and volatility-management studies use risk signals in ways that are not equivalent to cross-asset inverse-volatility allocation. Consequently, the literature reviewed here does not establish whether a more complex ETF rule adds value after controlling simultaneously for the underlying stock–bond risk choice, the return path, horizon, turnover, and performance objective. A rule may outperform 1/N but not a nearby fixed allocation, or improve the Sharpe ratio without maximizing wealth or certainty-equivalent utility. This study addresses that gap through a complete fixed-allocation grid, paired historical and bootstrap comparisons, multiple horizons, and joint reporting of economic magnitude, improvement probabilities, and multiplicity-adjusted significance. The research expectation is therefore conditional rather than directional.

3. Data

The analysis uses weekly adjusted closing prices for seven U.S.-listed ETFs from 2 January 2009 to 31 December 2025. After converting prices to returns, the usable sample contains 887 weekly observations from 9 January 2009. Adjusted prices incorporate distributions and corporate actions. The price series were compiled from publicly available Yahoo Finance data and the documentation of the respective asset-management companies (Yahoo Finance, 2026). The time-varying risk-free return is based on the three-month U.S. Treasury bill rate from the Federal Reserve Bank of St. Louis FRED database and is converted to a weekly rate (Federal Reserve Bank of St. Louis, 2026).

Table 1 reports the full source universe and the analytical status of each series. SPY, IJR, VGK, and EEM represent U.S. large-cap, U.S. small-cap, European, and emerging-market equities. LQD, TLT, and BIL represent investment-grade corporate bonds, long-term Treasury bonds, and short-term Treasury bills. GLD and VNQ were available in the source data but were excluded because the fixed rules are defined explicitly as stock–bond allocations. ACWI was retained only as an external market benchmark and is not counted among the fourteen strategies.

Table 1. ETF source universe and analytical status

Asset class	Ticker	Market segment	Status
Equities	SPY	U.S. large-cap equities	Included
Equities	IJR	U.S. small-cap equities	Included
Equities	VGK	European equities	Included
Equities	EEM	Emerging-market equities	Included
Bonds	LQD	U.S. investment-grade corporate bonds	Included
Bonds	TLT	Long-term U.S. Treasury bonds	Included
Bonds	BIL	Short-term U.S. Treasury bills	Included
Commodity	GLD	Gold	Excluded
Real estate	VNQ	U.S. real estate equities (REITs)	Excluded
Equities	ACWI	Global equities	Benchmark

Note. ETF = exchange-traded fund. "Included" denotes the seven-ETF opportunity set; "Excluded" denotes series available in the source data but omitted from the primary analysis; "Benchmark" denotes ACWI, which is used only as an external market comparator. Source: Author's compilation based on Yahoo Finance (2026), Investing.com (2026), and Stooq (2026).

Weekly simple returns are calculated as

$$R_{i,t} = \frac{P_{i,t}^{\text{adj}}}{P_{i,t-1}^{\text{adj}}} - 1 \tag{1}$$

where $R_{i,t}$ denotes the return on ETF i in week t and $P_{i,t}^{\text{adj}}$ is its adjusted closing price. Dates are sorted chronologically, duplicates are rejected, and all required series must be finite and complete. The same validated return matrix is used by every strategy.

4. Methods

4.1. Portfolio strategies and estimation

The eleven fixed strategies use target equity shares s in {0, 0.1, ..., 1}. The class weight is divided equally across the ETFs within that class:

$$w_i(s) = \begin{cases} \frac{s}{N_E}, & i \in E, \\ \frac{1-s}{N_B}, & i \in B, \end{cases} \qquad N_E = 4,\ N_B = 3 \tag{2}$$

where E and B denote the equity and bond subsets. The 1/N benchmark assigns

$$w_{i,t}^{1/N} = \frac{1}{N}, \qquad N = 7 \tag{3}$$

to each ETF. The 1/SD rule assigns inverse-volatility weights subject to a 30% cap:

$$w_{i,t}^{1/SD} = \min\left\{0.30, \frac{\lambda_t}{\widehat{\sigma}_{i,t}}\right\}, \qquad \sum_{i=1}^{N} w_{i,t}^{1/SD} = 1 \tag{4}$$

where *sigma-hat*$_{i,t}$ is the standard deviation of ETF i estimated from the preceding 156 weekly returns. The positive scaling constant *lambda*$_t$ is chosen so that the capped weights sum to one. This is equivalent to redistributing any excess above 30% proportionally across uncapped ETFs until the portfolio is fully invested.

The Markowitz strategy maximizes the estimated weekly return subject to the estimated variance of 1/N, full investment, long-only positions, and an upper bound on each ETF:

$$\max_{\mathbf{w}_t} \quad \mathbf{w}_t' \widehat{\boldsymbol{\mu}}_t \tag{5}$$

subject to

$$\mathbf{w}_t'\widehat{\boldsymbol{\Sigma}}_t\mathbf{w}_t \leq \left(\mathbf{w}^{1/N}\right)'\widehat{\boldsymbol{\Sigma}}_t\mathbf{w}^{1/N} \tag{6}$$

and

$$\sum_{i=1}^{N} w_{i,t} = 1, \qquad 0 \leq w_{i,t} \leq 0.40 \tag{7}$$

Here, $\hat{\mu_t}$ and $\hat{\Sigma_t}$ are the mean-return vector and covariance matrix estimated over the preceding 156 weeks. Thus, the maximum allocation to any single ETF in the Markowitz portfolio is 40%. A small ridge adjustment is added to the covariance matrix when required for numerical stability.

4.2. Out-of-sample design, rebalancing, and simulation

The first 156 weeks form the initial estimation window. Thereafter, the 1/SD and Markowitz parameters are re-estimated annually from a trailing 156-week window. Fixed and 1/N targets are also restored annually. Between rebalancing dates, weights drift with returns, and week-t performance uses weights determined before week t. The historical out-of-sample backtest consequently runs from 6 January 2012 to 31 December 2025.

One-way turnover at rebalancing date k is

$$TO_k = \frac{1}{2}\sum_{i=1}^{N}\left|w_{i,k}^{*} - w_{i,k}^{-}\right| \tag{8}$$

where $w_{i,k}^{*}$ is the new target and $w_{i,k}^{-}$ is the pre-trade weight. Initial formation is uncharged. With proportional cost c, the net return is

$$R_{p,t}^{net} = (1 - c\,TO_t)\left(1 + R_{p,t}^{gross}\right) - 1. \tag{9}$$

The baseline uses c = 0.001, or 10 basis points per unit of turnover; 0 and 25 basis points provide sensitivity tests.

Long-horizon outcomes are generated using 1,000 multivariate moving-block bootstrap paths with 13-week blocks (Künsch, 1989). Complete weekly return vectors are sampled jointly, preserving contemporaneous cross-asset dependence and short-run serial dependence. Every path contains 2,080 evaluated weeks. Outcomes at weeks 520, 1,040, 1,560, and 2,080 correspond to 10, 20, 30, and 40 years, so all horizons and all strategies share the same underlying simulated paths. Dynamic weights are re-estimated within each path using only the preceding simulated observations. A fixed random seed of 20260725 makes the experiment reproducible. All calculations were performed in R 4.6.0; the complete R script records the settings and produces the reported tables and figures.

4.3. Performance measures and statistical comparison

Portfolio wealth begins at V_0 = 100, with no subsequent deposits or withdrawals. For T weekly observations, CAGR and annualized volatility are

$$\mathrm{CAGR} = \left(\frac{V_T}{V_0}\right)^{52/T} - 1, \qquad \sigma_{\mathrm{ann}} = \mathrm{SD}\left(R_{p,t}\right)\sqrt{52}. \tag{10}$$

Risk-adjusted performance is measured by the Sharpe ratio (Sharpe, 1994) and the Sortino ratio (Sortino & Price, 1994):

$$\mathrm{SR} = \frac{\overline{R_{p,t} - R_{f,t}}}{\mathrm{SD}\left(R_{p,t} - R_{f,t}\right)}\sqrt{52}, \qquad \mathrm{Sortino} = \frac{\overline{R_{p,t} - R_{f,t}}}{\sqrt{T^{-1}\sum_{t=1}^{T}\min\left(R_{p,t} - R_{f,t}, 0\right)^2}}\sqrt{52}. \tag{11}$$

Maximum drawdown is (Chekhlov et al., 2005)

$$\mathrm{MDD} = \max_t\left[1 - \frac{V_t}{\max_{s \leq t} V_s}\right]. \tag{12}$$

The principal utility measure is the annualized certainty equivalent

$$\mathrm{CEQ}_j = 52\,\overline{R}_j - \frac{\gamma}{2}\,52\,\mathrm{Var}(R_j), \qquad \gamma = 3, \tag{13}$$

where γ is relative risk aversion. Weekly historical 95% Value at Risk and Expected Shortfall are retained as supplementary indicators, and average annual turnover equals cumulative one-way turnover divided by years.

Because every strategy faces the same path, comparisons with 1/N and 60/40 are paired:

$$\Delta M_{j,b,s} = M_{j,s} - M_{b,s}, \tag{14}$$

where $M_{j,s}$ is metric M for strategy j in simulation s and b denotes the benchmark. Reported statistics include the median paired difference, the 2.5th–97.5th percentile interval, and the probability of improvement. For loss measures, improvement means a lower value. Wilcoxon signed-rank tests assess a zero paired difference, and Holm's procedure controls family-wise error (Holm, 1979). Economic magnitude and improvement probability are interpreted together with adjusted significance.

5. Results

5.1. Historical out-of-sample performance

Table 2 summarizes the historical backtest at the baseline cost. The all-equity portfolio produces the highest terminal wealth, 385.55 from an initial 100, and the highest CAGR, 10.08%. Its 35.21% maximum drawdown, however, is materially larger than those of the balanced portfolios. Markowitz ends at 348.47 with a 9.29% CAGR, but delivers the highest Sharpe ratio (0.778), Sortino ratio (1.115), and CEQ (7.87%). Its maximum drawdown is 26.89%, compared with 23.85% for 1/N and 24.10% for 60/40.

Table 2. Historical out-of-sample results, 6 January 2012–31 December 2025

Strategy	CAGR	Vol.	SR	MDD	CEQ	Turn.
100/0	10.08	16.37	0.576	35.21	6.94	2.84
90/10	9.31	14.72	0.575	31.77	6.75	3.18
60/40	6.93	10.20	0.558	24.10	5.67	3.78
1/N	6.70	9.81	0.554	23.85	5.53	3.78
1/SD	4.98	7.03	0.508	19.50	4.37	4.34
Markowitz	9.29	10.11	0.778	26.89	7.87	18.35
0/100	1.86	6.81	0.078	23.49	1.37	2.12

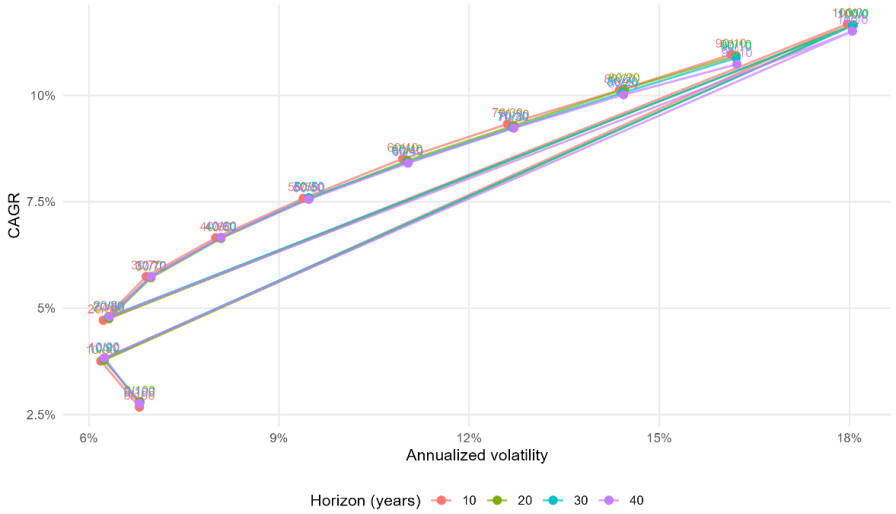
Note. Results use annual rebalancing and 10 basis points of transaction costs per unit of one-way turnover. CAGR, volatility, MDD, CEQ, and turnover are annualized or expressed in percent as applicable. Source: Author's calculations.

The historical 1/SD portfolio combines a 4.98% CAGR with 7.03% volatility and a 19.50% maximum drawdown. Its 4.37% CEQ remains below the 5.53% achieved by 1/N. The 30% cap limits concentration in BIL while preserving the strategy's lower-risk profile; historically, however, Markowitz remains strongest on risk-adjusted performance and the all-equity portfolio on wealth growth.

5.2. Baseline long-horizon simulations

The simulations preserve the historical trade-off between equity exposure and risk. Among fixed allocations, median CAGR rises almost monotonically from 2.68% for 0/100 to 11.68% for 100/0 at 10 years; the corresponding annualized volatilities are 6.80% and 17.97%. Similar values persist at longer horizons because the horizon affects compounding and estimation precision more strongly than the annualized first and second moments. Figure 1 maps the median risk–return profiles of the complete fixed-allocation grid.

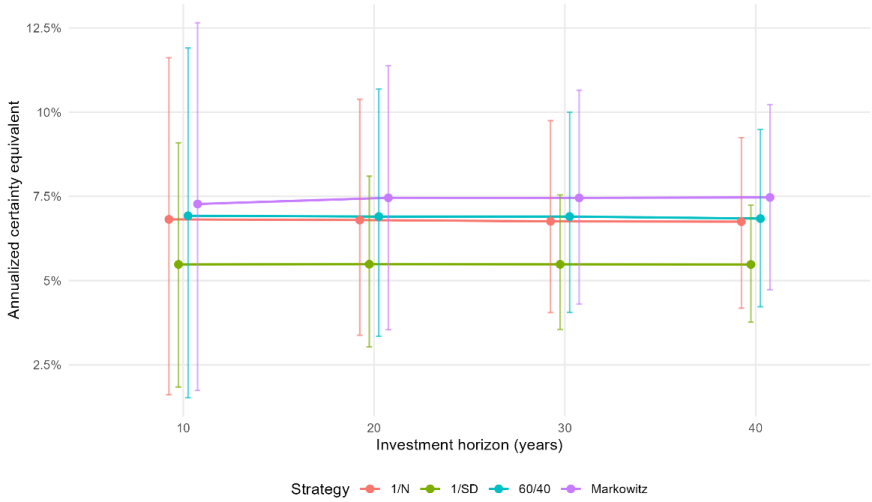
Fig 1. Fixed stock/bond allocations: simulated risk–return profiles.



Note. Points show median outcomes across 1,000 paired simulations; labels denote equity/bond target weights and transaction costs equal 10 basis points. Source: Author's calculations.

The ranking changes when performance is adjusted for risk. Markowitz has the highest median Sharpe ratio at every horizon, increasing from 0.722 at 10 years to approximately 0.736 at 20–40 years. It also produces the highest median Sortino ratio, between 1.060 and 1.081. Its median CEQ is 7.27–7.47%, compared with 6.75–6.82% for 1/N and 6.84–6.92% for 60/40. Among fixed allocations, 90/10 and 100/0 have economically indistinguishable 10-year median CEQs: the unrounded values are 7.745% and 7.742%, respectively, a difference of only 0.003 percentage points. The 100/0 allocation leads at 20, 30, and 40 years. Figure 2 emphasizes that the Markowitz utility advantage is meaningful relative to the balanced benchmarks but does not exceed every high-equity fixed allocation.

Fig 2. Annualized certainty equivalent for selected strategies.



Note. Points are medians and bars span the 5th–95th percentiles across 1,000 paired simulations; gamma = 3 and transaction costs equal 10 basis points. Source: Author's calculations.

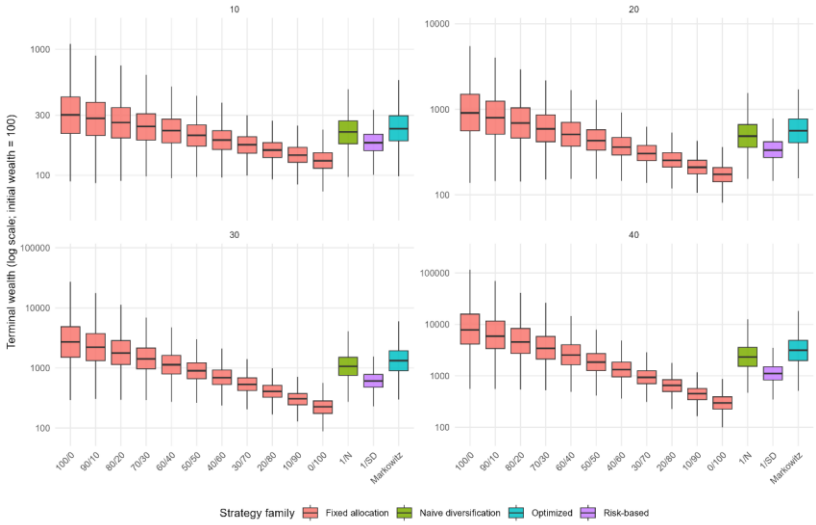
Table 3. Median simulated performance at 10 basis points

Years	Strategy	CAGR	Vol.	SR	MDD	CEQ
10	100/0	11.68	17.97	0.634	32.91	7.74
10	60/40	8.51	10.95	0.686	21.34	6.92
10	1/N	8.25	10.50	0.689	20.49	6.82
10	1/SD	6.14	7.21	0.695	14.65	5.48
10	Markowitz	8.89	10.72	0.722	20.06	7.27
20	100/0	11.66	18.06	0.639	36.90	7.83
20	60/40	8.47	11.02	0.680	23.63	6.90
20	1/N	8.22	10.57	0.684	22.79	6.80
20	1/SD	6.19	7.27	0.692	16.50	5.49
20	Markowitz	9.02	10.75	0.737	23.55	7.46
30	100/0	11.64	18.05	0.637	39.66	7.89
30	60/40	8.43	11.03	0.676	25.32	6.90
30	1/N	8.21	10.57	0.678	24.43	6.76
30	1/SD	6.20	7.28	0.686	17.51	5.48
30	Markowitz	9.00	10.76	0.735	24.91	7.46
40	100/0	11.51	18.05	0.624	42.12	7.65
40	60/40	8.41	11.04	0.668	26.61	6.84
40	1/N	8.18	10.57	0.671	25.67	6.75
40	1/SD	6.19	7.27	0.683	18.54	5.48
40	Markowitz	9.00	10.74	0.736	26.26	7.47

Note. Each entry is the median across 1,000 paired 13-week moving-block bootstrap paths. Vol. = annualized volatility; SR = Sharpe ratio; MDD = maximum drawdown; CEQ = annualized certainty equivalent. Source: Author's calculations.

Table 3 reports selected strategies. All-equity generates the largest median terminal wealth at every horizon: 301.80, 907.64, 2,719.03, and 7,806.68 after 10, 20, 30, and 40 years. The corresponding values are 234.38, 562.16, 1,326.85, and 3,138.21 for Markowitz; 220.97, 485.01, 1,066.52, and 2,320.03 for 1/N; and 181.42, 332.54, 608.44, and 1,106.31 for 1/SD. Figure 3 shows that wealth distributions widen markedly with horizon and equity exposure.

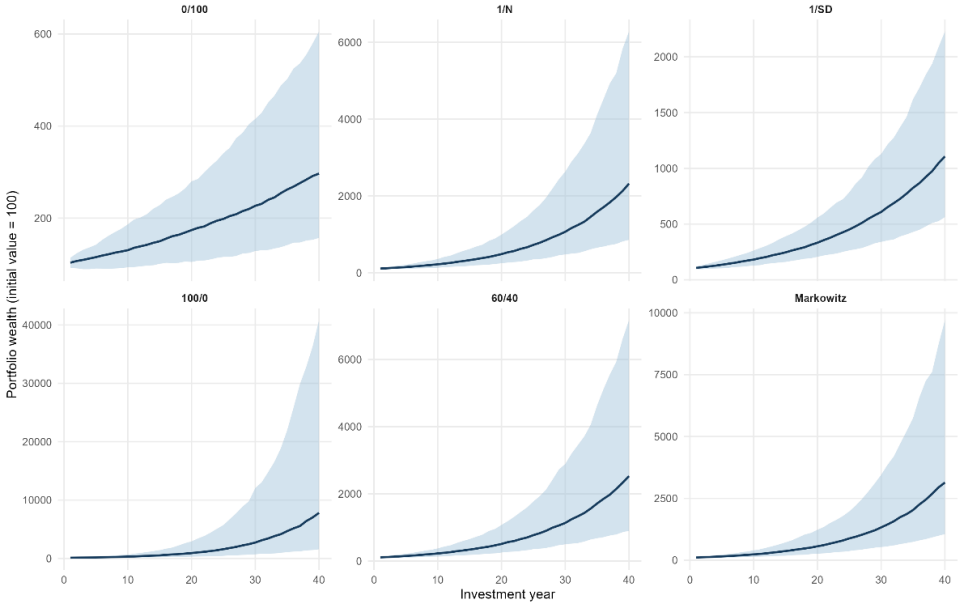
Fig 3. Distribution of terminal wealth across 1,000 paired simulations.



Note. Initial wealth equals 100; the vertical axis is logarithmic and transaction costs equal 10 basis points. Boxes show the interquartile range and median. Source: Author's calculations.

The 1/SD rule occupies a lower-risk middle ground after imposing the cap. Across the four horizons, its median CAGR is 6.14–6.20%, volatility is 7.21–7.28%, maximum drawdown is 14.65–18.54%, and CEQ is 5.48–5.49%. BIL remains at the 30% ceiling, but the remaining weight is diversified: median allocations are approximately 24% to LQD, 11% to TLT, and 35% in total across the four equity ETFs. Figure 4 shows that its 40-year wealth path lies between the bond-heavy and balanced rules.

Fig 4. Forty-year simulated wealth paths for selected strategies.



Note. The solid line is median wealth and the shaded band spans the 5th–95th percentiles across 1,000 paired simulations; initial wealth equals 100 and transaction costs equal 10 basis points. Source: Author's calculations.

5.3. Paired comparisons and transaction costs

Table 4 confirms that Markowitz improves risk-adjusted performance relative to both principal benchmarks. Against 1/N, the median paired CEQ gain ranges from 0.64 to 0.69 percentage points per year and the probability of a higher CEQ rises from 67.3% at 10 years to 82.6% at 40 years. The median Sharpe gain rises from 0.044 to 0.055, with improvement probabilities from 62.1% to 77.8%. Against 60/40, CEQ gains are 0.49–0.55 percentage points and improvement probabilities rise from 63.5% to 77.0%. All reported Markowitz differences in CAGR, Sharpe, and CEQ are significant after Holm adjustment at $p < 0.001$.

Table 4. Paired Markowitz differences relative to simple benchmarks

Years	Benchmark	ΔCAGR	Pr(CAGR+)	ΔSR	Pr(SR+)	ΔCEQ	Pr(CEQ+)
10	1/N	0.80	69.5	0.044	62.1	0.64	67.3
10	60/40	0.52	61.7	0.046	63.5	0.51	63.5
20	1/N	0.75	75.5	0.047	68.4	0.64	74.1
20	60/40	0.49	66.9	0.049	70.0	0.49	68.3
30	1/N	0.79	79.6	0.054	73.3	0.68	77.7
30	60/40	0.55	70.3	0.057	74.4	0.54	72.9
40	1/N	0.79	84.0	0.055	77.8	0.69	82.6
40	60/40	0.55	75.0	0.058	79.7	0.55	77.0

Note. Differences are paired medians across 1,000 common paths. change in CAGR and change in CEQ are percentage-point differences; Pr(+) is the probability that Markowitz has a higher value.

CAGR = compound annual growth rate; CEQ = annualized certainty equivalent. All Wilcoxon signed-rank tests shown have Holm-adjusted $p < 0.001$. Source: Author's calculations.

The advantage is not produced by a systematic increase in maximum drawdown. Relative to 1/N, median paired drawdown differences are only 0.42–0.72 percentage points; relative to 60/40, they range from −0.35 to −0.21 percentage points. Markowitz therefore has drawdown exposure close to the balanced benchmarks, even though its weights are more variable.

That variability is visible in turnover and target weights. Median annual turnover for Markowitz is 29.6–30.8%, compared with roughly 4.1–4.5% for 1/N and 4.5–4.9% for 1/SD. SPY frequently reaches the 40% constraint, while some regional equity positions often receive zero weight. Despite this activity, the cost sensitivity is modest: at a 10-basis-point trading cost, Markowitz loses approximately 3.2–3.4 basis points of annual CAGR; at 25 basis points, the drag is approximately 8.1–8.4 basis points. The CEQ ranking of the principal strategies is therefore stable across the tested costs.

6. Discussion

The findings reject both extreme interpretations of portfolio complexity. The constrained Markowitz rule consistently produces the highest median Sharpe and Sortino ratios and improves paired certainty-equivalent utility relative to 1/N and 60/40. It is not universally dominant, however: high-equity fixed allocations generate the greatest median terminal wealth; 90/10 and 100/0 are economically indistinguishable in 10-year median CEQ, and 100/0 leads at longer horizons. The 30%-capped 1/SD rule reduces risk without collapsing into a near-cash portfolio, but it still trails the balanced and optimized rules in median utility. Complexity therefore pays only relative to the investor's objective.

The Markowitz result is more favorable than the broad skepticism in DeMiguel et al. (2009), but the difference is economically interpretable. The opportunity set contains only seven diversified ETFs, the 156-week estimation window is long relative to the number of assets, short selling is prohibited, ex ante variance cannot exceed that of 1/N, and each ETF is capped at 40%. These restrictions regularize the optimizer, consistent with evidence that constraints can improve out-of-sample performance (Behr et al., 2013; Jagannathan & Ma, 2003) and with Kritzman et al. (2010). Frequent boundary solutions and materially higher turnover nevertheless support Michaud's (1989) warning that estimated portfolios remain sensitive to inputs.

The capped inverse-volatility results refine the evidence on parsimonious risk rules. BIL reaches the 30% limit, but the remaining weight is redistributed across bonds and equities; the median equity share is approximately 35%. Consequently, median volatility remains well below that of 1/N and 60/40, while median CAGR is approximately 6.2%. This is consistent with research showing that volatility information can alter portfolio risk efficiently (Kirby & Ostdiek, 2012; Moreira & Muir, 2017), but also with the caution that volatility management does not reliably improve out-of-sample utility (Cederburg et al., 2020). The cap therefore improves diversification, not dominance.

The fixed-allocation results clarify the role of horizon. High equity exposure compounds into a large median wealth advantage, but it also produces materially larger drawdowns and a wider distribution of outcomes. This is consistent with Anarkulova et al. (2022), who show that long horizons do not eliminate equity downside risk. Investors with limited loss capacity may therefore rationally accept lower median wealth in exchange for the narrower risk distribution delivered by a balanced, capped risk-based, or constrained optimized rule.

Transaction costs do not reverse the baseline rankings because rebalancing is annual. Even for Markowitz, the median annual CAGR drag is only about 3.2–3.4 basis points at the baseline cost and 8.1–8.4 basis points at 25 basis points. This accords with the broader evidence that rebalancing outcomes depend jointly on allocation, trading frequency, and

costs (Donohue & Yip, 2003; Dichtl et al., 2016). The result should not be generalized to taxable accounts, illiquid assets, or high-frequency optimization.

Several limitations remain. The 2009–2025 record is short relative to the simulated 40-year horizon and represents one U.S.-listed ETF universe and one currency environment. The block bootstrap preserves observed dependence but assumes historical blocks remain informative about future regimes. The analysis also uses nominal returns, one estimation window, annual rebalancing, a risk-aversion coefficient of three, no contributions or withdrawals, and proportional costs without taxes or market impact. Future research could examine shrinkage, robust optimization, asset-class-level risk weighting, or combinations of optimized and naïve weights such as Tu and Zhou (2011).

7. Conclusion

Within these limits, the conclusion is clear. Constrained optimization earns an economically meaningful and increasingly reliable risk-adjusted advantage as the horizon lengthens, while the capped inverse-volatility rule provides a transparent risk-reduction alternative. Simple high-equity portfolios remain superior for median wealth and for median certainty equivalent, although 90/10 and 100/0 are economically indistinguishable at 10 years. Portfolio complexity is therefore an objective-dependent tool: it can improve how risk is compensated, but it cannot determine how much risk the investor should bear.

Supplementary Materials: Please refer to any additional supplementary material available online: online appendices, datasets, codes, etc.

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, writing—original draft, writing—review and editing, and project administration: the sole author. The author has read and approved the manuscript.

Ethical Approval: This study did not involve human participants or any form of primary data collection. The analysis relied solely on secondary market data that were publicly available and contained no identifiable personal information. According to prevailing institutional and international guidelines, the use of publicly accessible secondary data does not require formal ethical approval. Nonetheless, the author adhered to all ethical principles governing the responsible use, citation, and reporting of secondary data in research.

Consent to Participate: This study did not involve direct interaction with human participants but relied exclusively on secondary market data available in the public domain. Accordingly, no individual-level consent to participate was required.

Funding: This research received no external funding.

Data Availability Statement: The analysis was implemented in R version 4.6.0 using the RStudio integrated development environment. The reproducibility R code and derived output tables are available from the author upon reasonable request. The underlying market data remain subject to the terms of their original providers.

Consent to Publish: The author has reviewed the final version of the manuscript and consents to its publication. The author affirms that the manuscript is original, has not been submitted elsewhere, and contains no material previously published without proper citation. The author accepts responsibility for the integrity and accuracy of the content and agrees to its submission for peer review and publication.

AI Use Statement: OpenAI ChatGPT was used for language editing and assistance in debugging the reproducible R code. The author verified all analyses, references, and interpretations and assumes full responsibility for the submitted content.

Acknowledgments: The author has no acknowledgments to report.

Declaration of Interest: The author declares no financial or non-financial interests, and no personal relationships, that could be construed as influencing the research presented in this paper or its outcomes.

References

Anarkulova, A., Cederburg, S., & O'Doherty, M. S. (2022). Stocks for the long run? Evidence from a broad sample of developed markets. *Journal of Financial Economics, 143*(1), 409–433. https://doi.org/10.1016/j.jfineco.2021.06.040

Behr, P., Guettler, A., & Miebs, F. (2013). On portfolio optimization: Imposing the right constraints. *Journal of Banking & Finance, 37*(4), 1232–1242. https://doi.org/10.1016/j.jbankfin.2012.11.020

Cederburg, S., O'Doherty, M. S., Wang, F., & Yan, X. (2020). On the performance of volatility-managed portfolios. *Journal of Financial Economics, 138*(1), 95–117. https://doi.org/10.1016/j.jfineco.2020.04.015

Chekhlov, A., Uryasev, S., & Zabarankin, M. (2005). Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance, 8*(1), 13–58. https://doi.org/10.1142/S0219024905002767

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies, 22*(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

Dichtl, H., Drobetz, W., & Wambach, M. (2016). Testing rebalancing strategies for stock–bond portfolios across different asset allocations. *Applied Economics, 48*(9), 772–788. https://doi.org/10.1080/00036846.2015.1088139

Donohue, C., & Yip, K. (2003). Optimal portfolio rebalancing with transaction costs. *The Journal of Portfolio Management, 29*(4), 49–63. https://doi.org/10.3905/jpm.2003.319894

Federal Reserve Bank of St. Louis. (2026). *3-month Treasury bill secondary market rate, discount basis [DTB3].* FRED. https://fred.stlouisfed.org/series/DTB3

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance, 58*(4), 1651–1683. https://doi.org/10.1111/1540-6261.00580

Kirby, C., & Ostdiek, B. (2012). It's all in the timing: Simple active portfolio strategies that outperform naïve diversification. *Journal of Financial and Quantitative Analysis, 47*(2), 437–467. https://doi.org/10.1017/S0022109012000117

Kritzman, M., Page, S., & Turkington, D. (2010). In defense of optimization: The fallacy of 1/N. *Financial Analysts Journal, 66*(2), 31–39. https://doi.org/10.2469/faj.v66.n2.6

Kunjal, D., & Peerbhai, F. (2026). Feedback trading in socially responsible ETFs: Does geopolitical risk matter? *Modern Finance, 4*(1), 24–35. https://doi.org/10.61351/mf.v4i1.520

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics, 17*(3), 1217–1241. https://doi.org/10.1214/aos/1176347265

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance, 10*(5), 603–621. https://doi.org/10.1016/S0927-5398(03)00007-0

Maillard, S., Roncalli, T., & Teïletche, J. (2010). The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management, 36*(4), 60–70. https://doi.org/10.3905/jpm.2010.36.4.060

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal, 45*(1), 31–42. https://doi.org/10.2469/faj.v45.n1.31

Moreira, A., & Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance, 72*(4), 1611–1644. https://doi.org/10.1111/jofi.12513

Page, D., Seetharam, Y., & Auret, C. (2026). Machine learning stock selection: Evidence from the South African factor zoo. *Modern Finance, 4*(3), 1–25. https://doi.org/10.61351/mf.v4i3.606

Pflug, G. C., Pichler, A., & Wozabal, D. (2012). The 1/N investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance, 36*(2), 410–417. https://doi.org/10.1016/j.jbankfin.2011.07.018

Poterba, J. M., & Shoven, J. B. (2002). Exchange-traded funds: A new investment option for taxable investors. *American Economic Review, 92*(2), 422–427. https://doi.org/10.1257/000282802320191732

Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management, 21*(1), 49–58. https://doi.org/10.3905/jpm.1994.409501

Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. *The Journal of Investing, 3*(3), 59–64. https://doi.org/10.3905/joi.3.3.59

Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics, 99*(1), 204–215. https://doi.org/10.1016/j.jfineco.2010.08.013

Yahoo Finance. (2026). *Historical market data.* https://finance.yahoo.com/

Disclaimer: All statements, viewpoints, and data featured in the publications are exclusively those of the individual author(s) and contributor(s), not of MFI and/or its editor(s). MFI and/or the editor(s) absolve themselves of any liability for harm to individuals or property that might arise from any concepts, methods, instructions, or products mentioned in the content.