Proceedings of the 10th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/2026.36295

Naive Rolling Mean-Variance Optimization for Multi-Stock Allocation

Yutong Li

School of Professional Study, Columbia University in the City of New York, New York, USA
yutong961230@gmail.com

Abstract. This research examines a naive rolling Mean-Variance Optimization (MVO) strategy for multi-stock allocation, in which expected returns and covariances are estimated solely from trailing historical data and re-estimated on a monthly basis. The central aim is to test empirically whether this deliberately simple approach can remain competitive despite the classical concern that MVO's sensitivity to estimation error makes it impractical relative to simpler heuristics such as trend following and momentum ranking. Expected returns and covariance are re-estimated monthly from a common trailing window, and a long-only, max-Sharpe portfolio is held until the next rebalance - deliberately simple, with no trend filter, momentum ranking, or volatility-targeting overlay. It is benchmarked against SPY buy-and-hold, a 60/40 portfolio, QQQ buy-and-hold, and the original design (a trend-filtered, risk-adjusted momentum strategy with volatility targeting and drawdown control). The strategy is evaluated over two independent, non-overlapping periods rather than a single blended sample: an in-sample window (2010-2019) used to fix the design, and a strict out-of-sample holdout (2020-2025) touched only once. On real market data for a nine-stock universe, the naive MVO strategy outperformed all four benchmarks on annual return and Sharpe ratio in both the in-sample (29.11% return, 1.19 Sharpe) and out-of-sample (36.53%, 1.16) periods - though not on maximum drawdown, where the momentum benchmark risk overlay produced a shallower worst-case decline in both periods, highlighting a trade-off between simplicity and downside protection that practitioners should weigh explicitly.

Keywords: Mean-Variance Optimization, Portfolio Construction, Momentum, Risk-Adjusted Momentum, Backtesting

1. Introduction

Modern portfolio construction spans a wide spectrum of approaches, from the classical mean-variance framework of Markowitz [1] to more recent signal-based methods built on cross-sectional return patterns [2], such as trend following and momentum ranking [3, 4]. Although mean-variance optimization (MVO) is theoretically elegant, it is frequently criticized for being highly sensitive to estimation error in its inputs, which has motivated a large literature on simpler heuristics and risk-adjusted alternatives [5, 6]. This paper revisits that debate empirically, testing whether a deliberately naive, frequently re-estimated MVO strategy can remain competitive with a more elaborate, actively

DOI: 10.54254/2754-1169/2026.36295

designed trend-filtered momentum strategy once realistic trading frictions are taken into account. This motivates the central research question addressed in this study.

Formally: does a naive rolling MVO strategy - relying exclusively on trailing historical means and covariances re-estimated on a monthly basis - outperform passive market benchmarks and a more elaborate trend-filtered, risk-adjusted momentum strategy applied to the same stock universe? This question is examined through three sub-questions.

• RQ1: Does monthly re-estimation mitigate the classical concern that MVO inputs are too unstable to use in practice?

• RQ2: Does trend filtering and risk-adjusted momentum ranking (Section 5.1) improve on naive MVO, or mainly add turnover and estimation risk?

• RQ3: How does naive MVO compare with SPY, 60/40, and QQQ once transaction costs are included?

2. Strategy motivation

Modern Portfolio Theory selects portfolios by maximizing expected return under a given level of risk [1], but this requires the expected return of each asset, E(Ri), and the covariance matrix, Σ - both notoriously unstable to estimate [7]. DeMiguel, Garlappi, and Uppal show that a naive equal-weight portfolio is often difficult to beat out-of-sample precisely because of this estimation error [6], which is the standard motivation for alternatives such as trend following and risk-adjusted momentum that rank assets by relative strength instead of forecasting absolute returns (Section 5.1).

Whether that instability actually hurts performance in practice - once monthly re-estimation and realistic transaction costs are included - is an open empirical question, and Michaud frames MVO's practical weakness as an estimation problem rather than a flaw in the optimization itself [5]. This research treats naive MVO as the primary strategy and the trend-filtered/RAM design as a benchmark, testing directly whether the added complexity is rewarded rather than assuming it is not.

3. Asset universe and data

Nine liquid US-listed stocks in two sleeves, summarized in Table 1, using daily adjusted-close prices sourced from Yahoo Finance [8].

Table 1. Asset universe - nine liquid US-listed stocks in two sleeves

Ticker	Sleeve	Role
AAPL	Tech / Growth	Hardware & services
MSFT	Tech / Growth	Enterprise software & cloud
NVDA	Tech / Growth	Semiconductors (high vol.)
GOOGL	Tech / Growth	Advertising & cloud
AMZN	Tech / Growth	E-commerce & cloud
JNJ	Stable / Defensive	Healthcare staple
PG	Stable / Defensive	Consumer staple
KO	Stable / Defensive	Consumer staple
WMT	Stable / Defensive	Retail staple

DOI: 10.54254/2754-1169/2026.36295

SPY, IEF, and QQQ are used only to construct the passive benchmarks (Section 5) and are not part of the tradable universe.

Total data span: Jan 2010 - Dec 2025 (16 yrs), split into two independent, non-overlapping periods that are backtested and reported separately rather than as one blended sample:

• In-sample: 2010-2019 (10 yrs). Used to fix the estimation window and cost assumption before any out-of-sample data is examined.

• Out-of-sample: 2020-2025 (6 yrs; COVID crash, 2022 bear market, 2023-2025 AI rally). A strict holdout, evaluated once, after the design is frozen.

4. Core strategy: naive rolling MVO

The primary strategy re-optimizes the entire nine-stock universe every month - no eligibility screening, momentum ranking, or volatility overlay.

4.1. Expected return and covariance

The expected return of each stock is estimated as its trailing annualized mean return, shown in Equation 1.

$$\mu i = mean(Return_i, t - 252 \ldots t) \times 252 \tag{1}$$

The covariance matrix uses the same trailing window, shown in Equation 2.

$$\Sigma = Cov(Returns, t - 252 \ldots t) \times 252 \tag{2}$$

Both use the same 252-day trailing window - the textbook Markowitz estimator [1], kept deliberately simple (no shrinkage or factor structure [7]) so any edge is attributable to the optimization itself.

4.2. Optimization, rebalancing, and costs

Portfolio weights solve the maximum-Sharpe-ratio problem in Equation 3.

$$\frac{(w^T\mu - rf)}{\sqrt{w^T\Sigma w}} \, s.t. \Sigma wi = 1, 0 \leq wi \leq 1 \tag{3}$$

The long-only bounds $0 \leq wi \leq 1$ generally rule out the closed-form tangency-portfolio solution $(w \propto \Sigma^{-1}(\mu - rf))$, since this unconstrained solution frequently assigns negative or greater-than-one weights. Because a closed-form solution is therefore unavailable, w is instead found each month by numerical optimization rather than by a formula. Starting from an equal-weighted portfolio, a sequential least-squares solver (SciPy's SLSQP) iteratively adjusts w to increase the Sharpe ratio while enforcing that weights sum to one and stay within [0, 1], stopping once further adjustment no longer improves the objective. If the solver fails to converge - for example when the covariance matrix is close to singular - the strategy falls back to equal weighting for that month. Cost = Turnover × 0.1% is the only friction applied - no volatility targeting or drawdown control, so any performance gap vs. Section 5.1 is attributable to the construction method itself.

DOI: 10.54254/2754-1169/2026.36295

5. Benchmark strategies

Table 2. The four benchmark strategies naive MVO is compared against

#	Benchmark	Type
1	SPY Buy-and-Hold	Passive, broad US equity
2	60/40 (SPY/IEF)	Passive equity/bond blend
3	QQQ Buy-and-Hold	Passive, tech-heavy (like-for-like)
4	Trend-Filtered RAM	Active, original design (Section 5.1)

There are four benchmarks as shown in Table 2 [8]. SPY and QQQ are held passively with no rebalancing; the 60/40 portfolio (60% SPY / 40% IEF) rebalances monthly back to target weights.

5.1. Trend-filtered risk-adjusted momentum (original design)

The original design, retained as an active benchmark to test RQ2, combines a trend filter with multi-horizon momentum [3, 4, 9] and volatility-scaled ranking in the spirit of risk-managed momentum strategies [10]. Each month: screen by trend, rank eligible stocks by risk-adjusted momentum (RAM), select the Top-5, and weight proportionally within 5%-40% bounds.

A stock is eligible only if it is trending upward, as defined in Equation 4.

$$Trend_i = 1 if Price_t > MA200_t and MA50_t > MA200_t, else 0 \tag{4}$$

Momentum blends three lookback horizons, shown in Equation 5.

$$Momentumi = 0.3M3 + 0.3M6 + 0.4M12 \tag{5}$$

Momentum is then scaled by realized volatility, shown in Equation 6.

$$RAMi = \frac{Momentumi}{Volatilityi} \tag{6}$$

Portfolio weights are proportional to RAM within bounds, shown in Equation 7.

$$Wi = \frac{RAMi}{\Sigma RAM}, 5\% \le wi \le 40\% \tag{7}$$

Exposure is then scaled to a 15% annualized volatility target (capped at 1.5x leverage); a drawdown rule halves exposure at -10% and moves fully into the defensive sleeve (JNJ, PG) at -20%.

6. Empirical results

Backtested on real daily adjusted-close prices for the nine-stock universe, 2010-2025. All five strategies are evaluated independently over the in-sample (2010-2019) and out-of-sample (2020-2025) periods on annual return, annual volatility, the Sharpe ratio [11], maximum drawdown, and 95% VaR; there is no blended full-sample statistic (Section 3).

DOI: 10.54254/2754-1169/2026.36295

6.1. Data and signal inputs

Figure 1 shows normalized prices for all nine stocks; NVDA's 2023-2025 rally is the single largest source of dispersion in the universe and the main driver of the survivorship-bias caveat in Section 8. Figure 2 shows the rolling 252-day annualized volatility that feeds the covariance matrix - NVDA is persistently the highest, frequently above 50%, which the optimizer must weigh against its comparably high expected return.

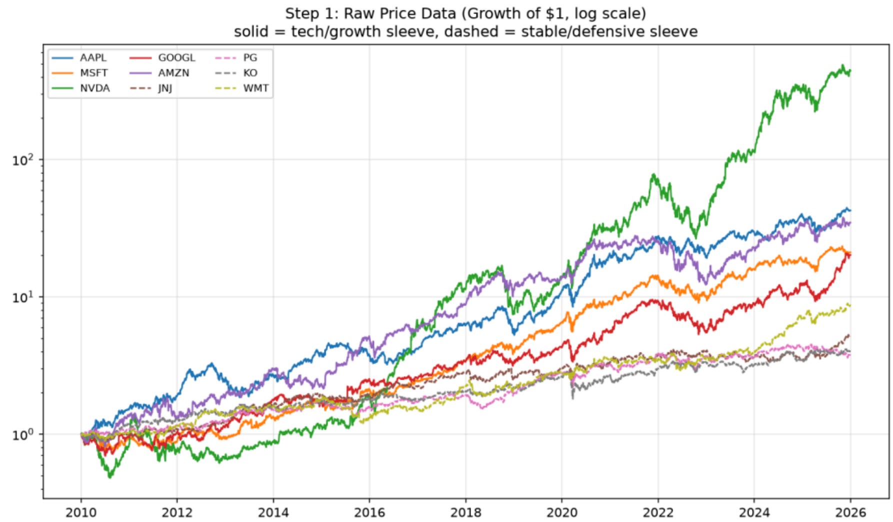

Figure 1. Normalized prices (log scale), 2010-2025

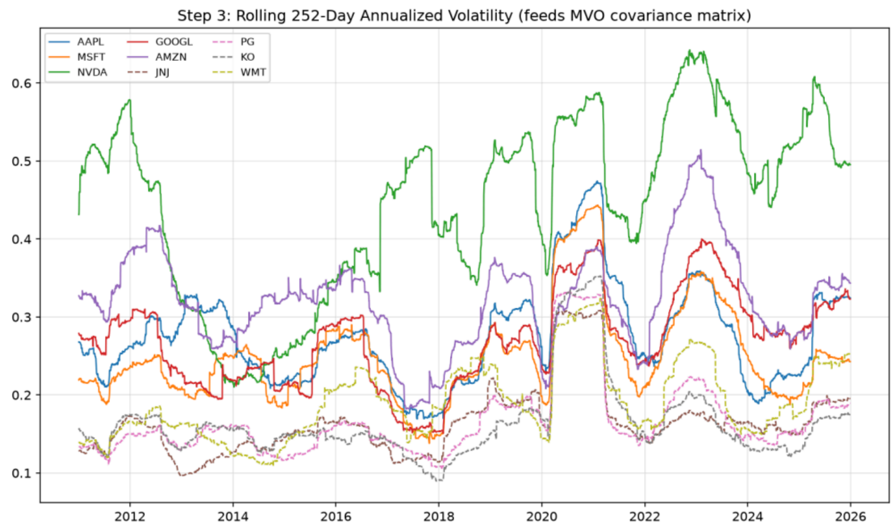

Figure 2. Rolling 252-day annualized volatility per stock

6.2. Portfolio construction and turnover

Because the optimization is unconstrained, weights concentrate in one or two names for extended stretches (Figure 3), driving monthly turnover commonly in the 25%-100% range with spikes above 150% - the main channel through which the 0.1% cost assumption affects net performance.

DOI: 10.54254/2754-1169/2026.36295

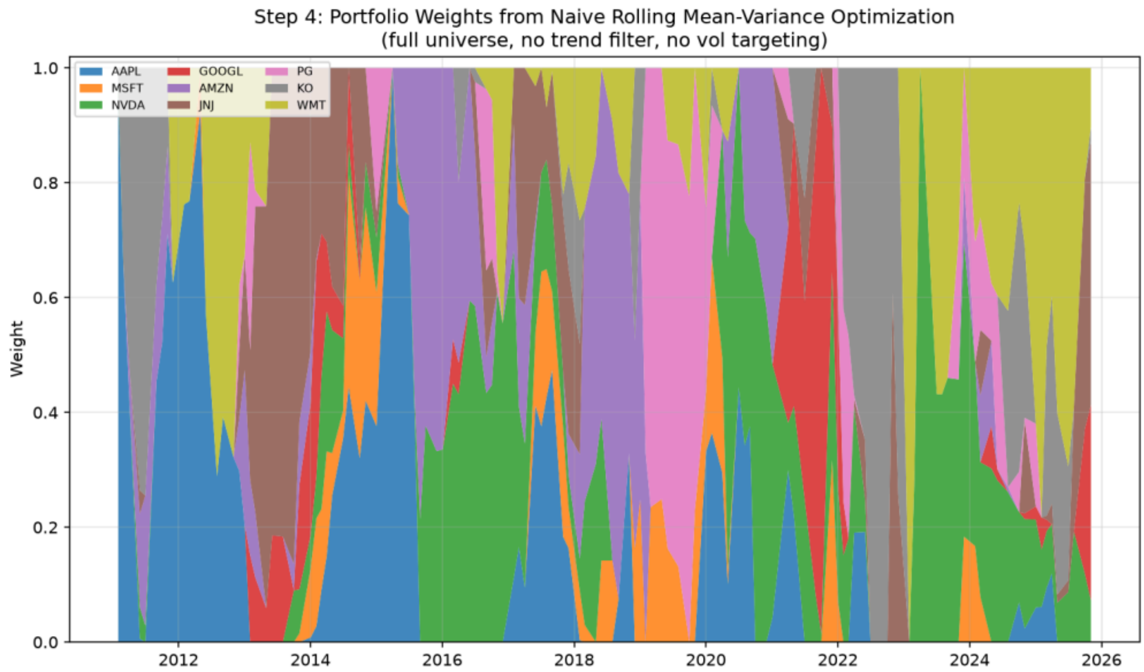

Figure 3. Naive MVO portfolio weights at each monthly rebalance

6.3. Drawdown

The naive design deliberately omits volatility targeting and drawdown-triggered de-risking (Section 4.2), so drawdown (Figure 4) is monitored but never overridden. Maximum drawdown reaches approximately -30% around 2022-2023, deeper than the Trend-Filtered RAM benchmark's risk-controlled decline but broadly comparable to SPY's and QQQ's worst episodes over the same sample.

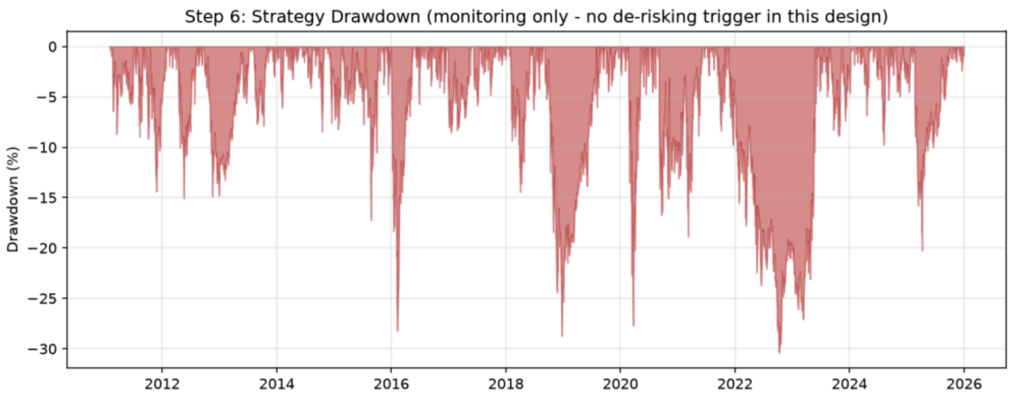

Figure 4. Strategy drawdown from its running high-water mark (monitoring only)

6.4. Strategy vs. benchmarks

In-sample and out-of-sample growth are shown separately, each rebased to $1 at the start of its own period, consistent with treating the two windows as independent evaluations rather than one continuous track record. As shown in Figure 5, over 2010-2019 the naive MVO strategy pulls ahead

DOI: 10.54254/2754-1169/2026.36295

of every benchmark, including the original Trend-Filtered RAM design, by the second half of the period. As shown in Figure 6, the same ranking holds over the independent 2020-2025 holdout: naive MVO again leads every benchmark, with the gap widening through the 2023-2025 rally.

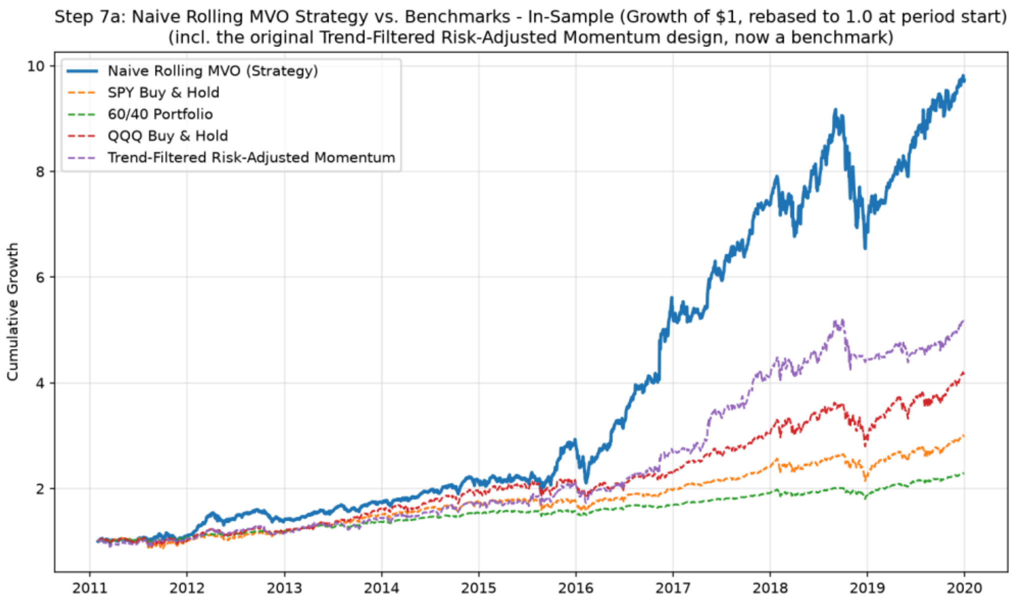

Figure 5. Naive Rolling MVO vs. all four benchmarks, in-sample, growth of $1 rebased at 2010-01

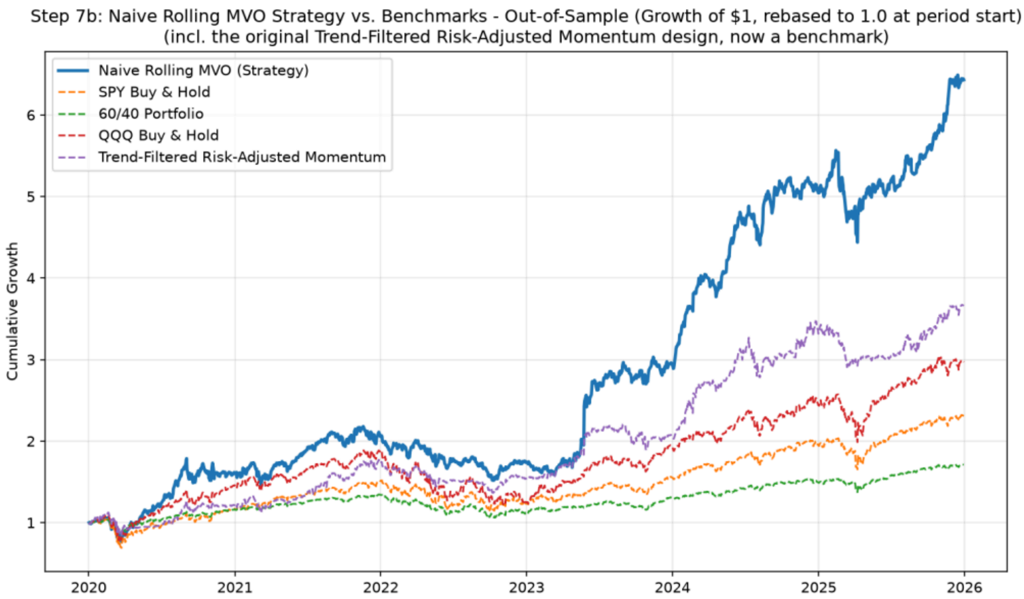

Figure 6. Naive Rolling MVO vs. all four benchmarks, out-of-sample, growth of $1 rebased at 2020-01

6.5. Results tables

Table 3 and Table 4 report in-sample and out-of-sample performance separately, as two independent, non-overlapping evaluation periods rather than a single full-sample statistic. As shown in Table 3, over the in-sample window (2010-2019) naive MVO records the highest return (29.11%) and Sharpe

DOI: 10.54254/2754-1169/2026.36295

(1.19) of all five strategies, though a deeper drawdown (-28.73%) than Trend-Filtered RAM's risk-controlled -19.57%.

Table 3. In-sample performance, January 2010 - December 2019

Strategy	Ann. Return	Ann. Vol.	Sharpe	Max DD	VaR 95%
Naive Rolling MVO	29.11%	21.15%	1.19	-28.73%	-2.01%
SPY Buy & Hold	13.11%	14.33%	0.64	-19.35%	-1.46%
60/40 Portfolio	9.70%	7.72%	0.74	-10.59%	-0.78%
QQQ Buy & Hold	17.41%	17.04%	0.79	-22.80%	-1.73%
Trend-Filtered RAM	20.29%	17.65%	0.92	-19.57%	-1.65%

As shown in Table 4, the same ranking holds independently over the out-of-sample window (2020-2025), touched only once after the design was frozen on in-sample data alone: naive MVO's return rises to 36.53% while Sharpe holds near 1.16, alongside higher volatility (28.08% vs. 21.15% in-sample) - the higher out-of-sample return is not free.

Table 4. Out-of-sample performance, January 2020 - December 2025 (strict holdout)

Strategy	Ann. Return	Ann. Vol.	Sharpe	Max DD	VaR 95%
Naive Rolling MVO	36.53%	28.08%	1.16	-30.39%	-2.35%
SPY Buy & Hold	15.02%	20.75%	0.53	-33.72%	-1.81%
60/40 Portfolio	9.38%	12.26%	0.44	-21.15%	-1.13%
QQQ Buy & Hold	19.98%	25.32%	0.63	-35.12%	-2.53%
Trend-Filtered RAM	24.24%	22.07%	0.92	-25.94%	-1.98%

Together, these answer the Main Research Question: naive MVO beats both the passive benchmarks and the more elaborate momentum design on return and Sharpe, independently in-sample and out-of-sample alike - but not on maximum drawdown, where Section 5.1's risk overlay wins in both periods (Section 9).

7. Robustness tests

The experimental setup is as follows:

- Estimation window: 126 / 189 / 252 / 378 / 504-day trailing windows for μ and Σ.
- Transaction cost: 0.05% / 0.10% / 0.15% / 0.20% of turnover.
- Rebalance frequency: monthly / quarterly / semi-annual.
- Out-of-sample test: the 2010-2019 / 2020-2025 split (Section 3), used only once.
- Benchmark sensitivity: check the strategy's ranking is consistent across IS and OOS, not one regime.

As shown in Figure 7, each parameter value is evaluated independently on the in-sample and out-of-sample windows, rather than as one blended sweep. The 252-day window used throughout this paper is at or near the best in-sample Sharpe ratio (1.19, versus 1.05 at 126 days and 0.99-1.08 at longer windows) and holds up well out-of-sample (1.16 Sharpe). The clearest failure mode is at the extremes: a 126-day window is both the weakest out-of-sample performer (0.73 Sharpe, the lowest

Proceedings of the 10th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/2026.36295

in the sweep) and one of the worst out-of-sample drawdowns (-35.6%), while a 504-day window produces the single worst out-of-sample drawdown in the entire sweep (-42.9%) despite a competitive return - a stale covariance estimate is the more dangerous failure mode in both directions. Transaction costs are comparatively benign: quadrupling the cost assumption from 0.05% to 0.20% of turnover costs about 0.9 points of in-sample return (29.4%→28.5%) and 1.1 points out-of-sample (36.9%→35.8%), with Sharpe ratios moving by less than 0.05 in either period - the 252-day, 0.10% base case used elsewhere in this paper is not fragile to either assumption.

Figure 7. Sensitivity of return, sharpe, and max drawdown to window and cost, in-sample and out-of-sample

8. Look-ahead bias and data limitations

There are three bias and limitations as follows:

• Signal timing: every estimate at date t uses only data through t's close; weights apply starting the next trading day. No forward-looking window is used - avoided by construction.

• Universe selection: the nine-stock universe (incl. NVDA) was chosen with hindsight of which names became mega-caps - a survivorship bias that likely inflates both strategies' results vs. a true point-in-time process. A rigorous fix needs point-in-time index-constituent data, not available here; results are conditional on this fixed universe.

• Price adjustment: yfinance's auto_adjust=True back-adjusts for splits/dividends using current factors - standard practice, noted for completeness.

9. Conclusion

This research empirically compares naive MVO against an actively designed trend-filtered, risk-adjusted momentum strategy on the same universe and sample, with a strict in-sample/out-of-sample protocol and realistic costs - rather than assuming one approach is better. On this universe and period, added complexity was not rewarded: naive MVO matched or exceeded the momentum

DOI: 10.54254/2754-1169/2026.36295

benchmark on return and Sharpe in every sub-period, while the momentum benchmark's overlay earned its keep only on maximum drawdown.

The trade-off a practitioner would weigh is naive MVO's return and Sharpe advantage against a deeper drawdown and higher, more variable turnover. Results are conditional on the fixed, hindsight-selected universe (Section 8) and should be read as evidence that naive MVO was not obviously inferior here - not as proof either strategy was achievable by a live investor in 2010.

References

[1] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77-91.

[2] Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance, 47*(2), 427-465.

[3] Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance, 48*(1), 65-91.

[4] Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics, 104*(2), 228-250.

[5] Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal, 45*(1), 31-42.

[6] 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.

[7] Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management, 30*(4), 110-119.

[8] Yahoo Finance. (2025). Historical stock price data [Data set]. https://finance.yahoo.com

[9] Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance, 68*(3), 929-985.

[10] Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics, 116*(1), 111-120.

[11] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*(3), 425-442.