



The Impact of Artificial Intelligence on Portfolio Performance: A Comparative Study of Machine Learning and Traditional Portfolio Optimisation Approaches

Shaik Reshma¹, Mr. V. Narendra²

¹ Research Scholar, KLBS, Koneru Lakshmaiah Education Foundation (KLEF)

² Assistant Professor, KLBS, Koneru Lakshmaiah Education Foundation (KLEF)

DOI : <https://doi.org/10.55248/gengpi.07.1026.3296>

ABSTRACT :

Artificial intelligence and machine learning are now being used in investment products, but it remains uncertain whether these tools consistently outperform established portfolio optimisation methods or control risk more effectively. This study compares three AI/ML-oriented exchange-traded funds—AIEQ, QRFT and AMOM—with three funds following minimum-variance or low-volatility strategies—USMV, SPLV and ACWV. The comparison covers 2021–2025 and relies on secondary data for annual total returns, the S&P 500 total-return benchmark and the three-month Treasury bill rate. Performance is assessed using returns, Sharpe and Sortino ratios, and the worst annual loss. Risk-management efficiency is represented by a combined measure based on volatility, downside capture, worst-year loss and beta. The analysis uses chi-square tests, one-way and two-way ANOVA, multiple regression and mediation testing. Traditional funds recorded stronger risk-management efficiency ($F = 67.9$, $p = 0.001$), while the difference in average annual returns between the groups was not statistically significant (10.8% for ML-based funds and 7.8% for traditional funds; $p = 0.622$). Although the regression model was significant overall ($R^2 = 0.76$), market conditions were the only significant individual predictor. Results also indicate that the relative performance of the two approaches changes with the market: ML-based funds did better in rising periods but experienced substantially larger losses in 2022. Risk-management efficiency did not significantly explain returns, and the proposed mediation effect was not supported. Given the limited sample, these findings should be treated as preliminary rather than conclusive.

Keywords: artificial intelligence, machine learning, portfolio optimisation, minimum variance, risk management efficiency, Sharpe ratio, portfolio performance.

Introduction

One of the main decisions in investing is how to distribute money across different assets. Markowitz (1952) established the importance of evaluating risk and expected return together, making mean–variance optimisation a core part of modern portfolio theory. Other approaches, including minimum-variance portfolios, Sharpe-ratio optimisation and the Black–Litterman model, later developed this foundation. However, these methods rely on estimates of future returns and relationships between assets. When those estimates are inaccurate, the resulting portfolio weights may become unstable and performance can suffer (Michaud, 1989; DeMiguel et al., 2009).

Machine learning provides another way to analyse financial information. Techniques such as random forests, neural networks, long short-term memory (LSTM) networks and reinforcement learning can identify complex, non-linear relationships in large datasets and can be updated as new information becomes available (Gu et al., 2020; Heaton et al., 2016). Researchers have begun using these tools either to improve the forecasts supplied to conventional optimisers or to determine portfolio weights directly. However, it is still an open question whether these methods deliver consistently stronger results and whether any improvement comes from more effective risk control.

This research compares the two approaches using publicly available data for the five-year period from 2021 to 2025. It also examines whether differences in risk-management efficiency help explain any differences in investment performance.

Statement of the Problem

AI-based investment products are becoming more visible to investors and fund managers, but research has not reached a clear conclusion about whether they outperform established optimisation methods. Some evidence comes from simulated back-tests, narrow asset classes such as cryptocurrency, or periods that do not include both strong and weak markets (Jiang et al., 2017; Betancourt & Chen, 2021). Other research suggests that even advanced optimisation methods may fail to outperform simple diversification strategies when estimation errors are considered (DeMiguel et al., 2009). A further question is whether AI-oriented funds control risk more effectively and whether that risk control translates into better returns. This study explores these

issues by comparing the actual performance of six listed funds over five years.

Review of Literature

Modern portfolio theory begins with Markowitz (1952), who set out a framework for selecting portfolios by balancing expected return against variance and established the efficient frontier. Its practical weakness was soon recognised. Michaud (1989) argued that mean-variance optimisers amplify errors in their inputs and can end up with portfolios that are inferior to simple equal weighting. Black and Litterman (1992) responded by combining equilibrium returns with investor views to obtain more stable weights, and Ledoit and Wolf (2004) showed that shrinking the sample covariance matrix reduces estimation error and improves performance. Even so, DeMiguel et al. (2009) tested fourteen optimisation models on seven data sets and found that none consistently beat the naive 1/N portfolio out of sample, which reinforced doubts about the value of conventional optimisation.

The arrival of ML in finance shifted attention towards better prediction. Heaton et al. (2016) discussed how deep learning can capture complex interactions in financial data that conventional models miss. Gu et al. (2020) compared a wide range of ML methods for predicting asset risk premiums and found that trees and neural networks produced large economic gains because they capture non-linear interactions among predictors. Related work by Kelly et al. (2019), Freyberger et al. (2020) and Kozak et al. (2020) used latent-factor, non-parametric and shrinkage methods to handle large sets of firm characteristics, although these papers are mainly about asset pricing and do not test portfolio performance directly.

A second group of studies links prediction with portfolio construction. Wang et al. (2020) combined LSTM return forecasts with mean-variance optimisation and reported better return and risk measures than five benchmark strategies, and Ma et al. (2021) found that a mean-variance model fed with random forest forecasts performed best among the approaches they tested. Zhang et al. (2020) went a step further by training a deep network to maximise the Sharpe ratio directly, and Jang and Seong (2023) connected deep reinforcement learning with modern portfolio theory, improving the Sharpe ratio, annual return and drawdown relative to benchmarks. Reinforcement learning has also been applied to dynamic allocation, including markets with a changing number of assets (Jiang et al., 2017; Betancourt & Chen, 2021; Gollart & Okhrin, 2025).

Comparative evidence is more recent. Ngo et al. (2023) compared reinforcement learning, deep learning and traditional optimisation in Vietnamese and U.S. stock markets and found that reinforcement learning performed best and conventional optimisation worst. Ha et al. (2026) tested several ML algorithms against the traditional All-Weather portfolio using daily ETF data, and Babiak and Baruník (2026) found that deep-learning forecasts improved portfolio choice, with the gains concentrated in recessions and large market movements. Simon et al. (2026) showed that deep parametric portfolio policies earned higher certainty-equivalent returns than comparable linear policies, and Wang et al. (2026) argued that integrating prediction and optimisation in a single end-to-end process outperforms the usual two-stage approach. On the risk side, Agal et al. (2025) proposed an ML architecture that combines volatility forecasting, risk budgeting and regime switching, and López de Prado et al. (2025) discussed how ML can strengthen alpha generation and risk management within the Markowitz paradigm.

Taken together, the literature suggests that ML can improve both prediction and portfolio construction, that its benefits may be larger in turbulent markets, and that risk management is one of the channels through which it may work. At the same time, many of the results come from the authors' own experiments, from particular asset classes or from working papers that have not yet been replicated.

Research Gap

Four main gaps emerge from the literature. First, many studies compare models with researcher-selected benchmarks rather than assessing established investment products using their observed returns. Second, the evidence often focuses on a small number of ETFs, one market or assets such as cryptocurrencies, without clearly comparing different market environments. Third, risk is commonly described using measures such as volatility or drawdown, but fewer studies test whether risk management actually explains the relationship between an investment approach and its results. Finally, relatively few studies consider the optimisation approach, risk-management efficiency and market conditions together in one analytical framework. The present study addresses these gaps by examining five years of secondary data across different market conditions.

Conceptual Framework

The framework adopted in this study, shown in Figure 1, links three sets of variables. The independent variable is the portfolio optimisation approach, which is either machine learning based (ML models, predictive analytics and AI-driven optimisation) or traditional (mean-variance, minimum-variance and Sharpe-ratio optimisation).

The mediating variable is risk management efficiency, reflecting how well a portfolio controls risk. It is described through risk-adjusted return, downside risk, diversification and volatility control. The idea is that an approach influences performance largely through the quality of its risk management.

The dependent variable is portfolio performance, measured by portfolio return, Sharpe ratio, volatility, maximum drawdown and Sortino ratio. Market conditions (bull, sideways or bear) are included as a control variable in the regression model, because the relative merit of an approach may differ between rising and falling markets (Babiak & Baruník, 2026).

Conceptual Framework

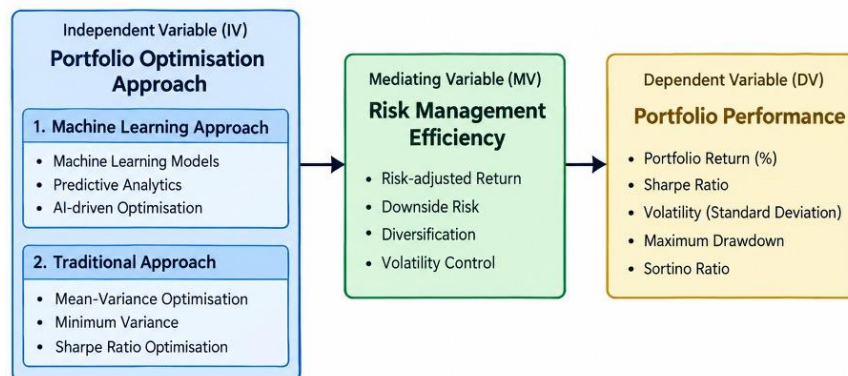


Figure 1: Conceptual Framework of the Study

Objectives of the Study

- To develop a conceptual framework linking portfolio optimisation approach, risk management efficiency, market conditions and portfolio performance.
- To examine the effect of Machine Learning and traditional portfolio optimisation approaches on risk management efficiency.
- To compare the portfolio performance of Machine Learning-based optimisation approaches with traditional portfolio optimisation approaches.
- To determine the effect of risk management efficiency on portfolio performance.

Hypotheses of the Study

- **H01:** Portfolio optimisation approach, risk management efficiency and market conditions have no significant relationship with portfolio performance.
- **H11:** Portfolio optimisation approach, risk management efficiency and market conditions have a significant relationship with portfolio performance.
- **H02:** There is no significant difference in risk management efficiency between Machine Learning-based and traditional portfolio optimisation approaches.
- **H12:** There is a significant difference in risk management efficiency between Machine Learning-based and traditional portfolio optimisation approaches.
- **H03:** There is no significant difference in portfolio performance between Machine Learning-based and traditional portfolio optimisation approaches.
- **H13:** There is a significant difference in portfolio performance between Machine Learning-based and traditional portfolio optimisation approaches.
- **H04:** Risk management efficiency has no significant effect on portfolio performance.
- **H14:** Risk management efficiency has a significant effect on portfolio performance.

Research Methodology

Research design

The research uses a descriptive, comparative and analytical design. It outlines the returns and risk profiles of AI/ML-driven funds and traditional optimisation funds, compares their results, and tests the relationships proposed in the conceptual framework. The analysis is based entirely on secondary information; no primary survey data or respondents are involved.

Sample and period

The sample consists of six exchange-traded funds whose portfolios are built by a stated approach. Three are classified as ML-based because their holdings are chosen by an artificial intelligence system: the Amplify AI Powered Equity ETF (AIEQ), which uses a model run on the IBM Watson platform, the QRAFT AI-Enhanced U.S. Large Cap ETF (QRFT) and the QRAFT AI-Enhanced U.S. Large Cap Momentum ETF (AMOM). Three are classified as traditional because they track indices constructed by minimum-variance optimisation or low-volatility selection: the iShares MSCI USA Min Vol Factor ETF (USMV), the iShares MSCI Global Min Vol Factor ETF (ACWV) and the Invesco S&P 500 Low Volatility ETF (SPLV). The S&P 500 Total Return Index serves as the market benchmark. Funds were chosen purposively because each has a published methodology and a

complete return history for the period. The period of study is the five calendar years 2021 to 2025, which contain a strong rising market (2021, 2023 and 2024), a moderate one (2025) and one sharp decline (2022). This gives 30 fund-year observations.

Table 1: Funds Included in the Study

Ticker	Fund	Group	Basis of classification
AIEQ	Amplify AI Powered Equity ETF	ML-based	AI model (IBM Watson) selects stocks
QRFT	QRAFT AI-Enhanced U.S. Large Cap ETF	ML-based	AI-based security selection
AMOM	QRAFT AI-Enhanced U.S. Large Cap Momentum ETF	ML-based	AI-based selection with momentum factor
USMV	iShares MSCI USA Min Vol Factor ETF	Traditional	Minimum-variance optimisation (MSCI)
ACWV	iShares MSCI Global Min Vol Factor ETF	Traditional	Minimum-variance optimisation (MSCI)
SPLV	Invesco S&P 500 Low Volatility ETF	Traditional	Low-volatility index selection

Source: Fund prospectuses and fact sheets.

Data source

Annual total-return figures were gathered from the public sources identified in the original dataset: Yahoo Finance for AIEQ, QRFT and AMOM; iShares fact sheets dated 30 June 2026 and Yahoo Finance for USMV and ACWV; and YCharts for SPLV. The S&P 500 total-return benchmark was also obtained from YCharts. The risk-free rate is the annual average of the Federal Reserve's three-month Treasury bill secondary-market rate, available through FRED. Fund returns use market prices and assume that distributions are reinvested. The accompanying workbook records the figures and their sources. The annual risk-free rates for 2021–2025 were 0.05%, 2.02%, 5.07%, 4.97% and 4.07%, respectively.

Variables and measures

Table 2: Variables and Their Measurement

Variable	Role	Measurement
Optimisation approach	Independent	ML-based = 1, traditional = 0 (dummy)
Risk management efficiency (RME)	Mediating	Index of four risk measures per fund over the five years: low standard deviation of annual returns, low downside capture (fund return divided by S&P 500 return in 2022), small worst-year loss and low beta. Each measure is converted to a z-score and the four are averaged
Market condition	Control	Annual S&P 500 total return; classed as bear (below 0%), moderate bull (0 to 20%) or strong bull (above 20%)
Portfolio return	Dependent	Annual total return (%)
Active return	Dependent	Fund return minus S&P 500 return (percentage points)
Sharpe ratio	Dependent	Mean annual excess return over the T-bill rate divided by its standard deviation
Sortino ratio	Dependent	Mean annual excess return divided by downside deviation
Worst-year loss	Dependent	Lowest annual return in the period (an annual proxy for maximum drawdown)

Source: Researcher's construction based on the conceptual framework.

Since the dataset contains annual rather than daily observations, the risk index reflects variation in yearly returns and each fund's behaviour during the 2022 downturn; it does not capture day-to-day volatility. The Sharpe and Sortino ratios are excluded from the risk-management-efficiency (RME) index to avoid building performance measures directly into the risk index.

Tools of analysis

- Descriptive statistics, fund-level risk measures and ratio analysis.
- Chi-square test of independence between approach and (a) beating the S&P 500, (b) earning more than the risk-free rate and (c) earning more than the sample median. Because some expected frequencies are below 5, Fisher's exact test is also reported. Cramér's V gives the effect size.
- One-way ANOVA to compare return and active return across approaches and across market conditions on the 30 fund-years, and to compare RME and the performance measures across approaches on the six funds (H02 and H03). Welch's t-test and the Mann-Whitney test are reported as checks, and Levene's test checks equal variances.
- Two-way ANOVA with approach and market condition as factors, including their interaction.

- Multiple linear regression: $\text{Return} = b_0 + b_1 (\text{Approach}) + b_2 (\text{RME}) + b_3 (\text{Market condition}) + e$, used to test H01, with variance inflation factors and the Durbin-Watson statistic as diagnostics.
- Simple regression of performance on RME to test H04, and a mediation analysis with the Baron and Kenny (1986) steps and the Sobel (1982) test.

A significance level of 5 per cent is used. The calculations can be repeated with the Python script and Excel workbook supplied with this paper.

Scope

The six funds are not a random sample, and the results describe these funds over these five years. With only six funds and five years, the fund-level tests have low power and the fund-year observations are not independent, because each fund contributes five. P-values are therefore indicative.

Data Presentation

Table 3: Calendar-Year Total Returns of the Sample Funds (%)

Fund	Group	2021	2022	2023	2024	2025
AIEQ	ML-based	20.10	-31.90	26.47	12.54	13.96
QRFT	ML-based	22.74	-22.70	24.16	21.36	17.95
AMOM	ML-based	13.09	-26.30	27.05	35.79	7.69
USMV	Traditional	20.84	-9.42	10.34	15.75	7.66
SPLV	Traditional	24.13	-4.89	0.53	13.92	4.10
ACWV	Traditional	13.95	-10.34	8.26	11.39	11.05
S&P 500 TR	Benchmark	28.71	-18.11	26.29	25.02	17.88
3-month T-bill	Risk-free	0.05	2.02	5.07	4.97	4.07

Source: Yahoo Finance, iShares, YCharts and FRED, as described under data source.

Data Analysis and Interpretation

Fund-level performance and risk

Table 4: Performance and Risk Measures of Each Fund, 2021 to 2025

Fund	Mean return (%)	CAGR (%)	SD (%)	Sharpe	Sortino	Worst year (%)	Beta	Downside capture	RME
AIEQ	8.2	5.8	23.1	0.22	0.33	-31.9	1.16	1.76	-1.19
QRFT	12.7	11.0	19.9	0.49	0.86	-22.7	1.02	1.25	-0.53
AMOM	11.5	9.1	23.9	0.36	0.65	-26.3	1.11	1.45	-0.92
USMV	9.0	8.5	11.5	0.49	1.13	-9.4	0.56	0.52	0.74
SPLV	7.6	7.1	11.5	0.34	1.17	-4.9	0.41	0.27	1.07
ACWV	6.9	6.5	9.8	0.37	0.66	-10.3	0.49	0.57	0.82
S&P 500 TR	16.0	14.4	19.5	0.67	1.41	-18.1	1.00	1.00	n/a

Source: Calculated from Table 3. Beta and capture are measured against the S&P 500 using five annual observations. SD = standard deviation of annual returns.

The fund-level results point to two distinct patterns. The three ML-based funds earned a higher average annual return (10.8 per cent against 7.8 per cent), but they were about twice as volatile, with a standard deviation of annual returns of 22.3 per cent against 10.9 per cent. Their betas were above one (1.10 on average) while the traditional funds had betas of about 0.48. In 2022, the ML-based funds lost on average 27.0 per cent in their worst year compared with 8.2 per cent for the traditional funds, and they fell about 1.49 times as much as the S&P 500, whereas the traditional funds fell only 0.45 times as much. After adjusting for risk the two groups were close: the average Sharpe ratio was 0.36 for ML-based and 0.40 for traditional funds, and the Sortino ratio, which looks only at downside risk, was higher for the traditional group (0.99 against 0.61). None of the six funds matched the S&P 500, whose Sharpe ratio was 0.67 and whose compound return was 14.4 per cent a year. On average, ML-based funds trailed the S&P 500 by 5.2 percentage points a year ($p = 0.011$) and traditional funds by 8.1 points ($p = 0.010$).

Chi-square test**Table 5: Chi-Square Tests of Independence (n = 30 fund-years)**

Cross-tabulation	Chi-square	df	p-value	Fisher p	Cramér's V
Approach x Beat S&P 500	0.19	1	0.666	1.000	0.08
Approach x Return above risk-free	0.19	1	0.666	1.000	0.08
Approach x Return above sample median	3.33	1	0.068	0.143	0.33

Source: Calculated from the fund-year panel. In the first two tests the smallest expected frequency is 3.5, so Fisher's exact p-value is the safer figure.

Neither the first nor the second test is significant. ML-based funds beat the S&P 500 in 4 of their 15 fund-years and traditional funds in 3 of 15, so the choice of approach had no bearing on whether a fund beat the market, and both groups mostly failed to do so. Likewise, 12 ML-based and 11 traditional fund-years earned more than the risk-free rate. The third test, whether a fund-year return was above the sample median, comes closest to significance: 10 of 15 ML-based fund-years were above the median against 5 of 15 for traditional funds (chi-square = 3.33, $p = 0.068$, Fisher $p = 0.143$). The association is moderate in size ($V = 0.33$) but not significant at 5 per cent. A test of approach against market condition was not run, as every fund is observed in every year and the table would be balanced by design.

Analysis of variance**Table 6: One-Way ANOVA on Fund-Years (n = 30)**

Dependent variable and factor	F	df	p-value	Eta-squared	Levene p
Return by Approach	0.25	1, 28	0.622	0.01	0.178
Active by Approach	0.83	1, 28	0.369	0.03	0.276
Return by Market	40.40	2, 27	<0.001	0.75	0.088

Source: Calculated from the fund-year panel.

Table 7: One-Way ANOVA on Fund-Level Measures, ML-Based versus Traditional (n = 6)

Measure	F (1, 4)	p-value	Welch p	Mann-Whitney p	Eta-squared
Risk management efficiency (H02)	67.87	0.001	0.004	0.10	0.94
Standard deviation	72.68	0.001	0.004	0.10	0.95
Worst-year loss	35.15	0.004	0.007	0.10	0.90
Downside capture	35.15	0.004	0.007	0.10	0.90
Mean annual return (H03)	4.07	0.114	0.141	0.20	0.50
Sharpe ratio (H03)	0.26	0.635	0.641	0.70	0.06
Sortino ratio (H03)	2.76	0.172	0.173	0.20	0.41

Source: Calculated from Table 4. With three funds per group, 0.10 is the smallest p-value the Mann-Whitney test can give.

The ANOVA using fund-year observations does not show a statistically significant difference in annual return between the approaches ($F = 0.25$, $p = 0.622$), nor in active return against the S&P 500 ($p = 0.369$). By contrast, market condition explains about 75 per cent of the variation in returns ($F = 40.4$, $p < 0.001$). At the fund level, risk management efficiency differs sharply between the groups: traditional funds scored 0.88 and ML-based funds - 0.88 ($F = 67.9$, $p = 0.001$, Welch $p = 0.004$). The same holds for the standard deviation, worst-year loss and downside capture. The Mann-Whitney test cannot go below 0.10 with three funds in each group, so the significant parametric result should be seen as supportive and not definitive. The Sharpe ratio ($p = 0.635$), Sortino ratio ($p = 0.172$) and mean return ($p = 0.114$) do not differ significantly.

Table 8: Two-Way ANOVA, Approach and Market Condition (Dependent Variable: Annual Return)

Source	F	df1	df2	p-value
Approach	13.61	1	24	0.001

Source	F	df1	df2	p-value
Market condition	13.36	2	24	<0.001
Approach x Market	11.63	2	24	<0.001

Source: Calculated from the fund-year panel.

Table 9: Mean Annual Return (%) by Approach and Market Condition

Market condition (S&P 500 return)	ML-based	Traditional	Difference
Strong bull (2021, 2023, 2024)	22.6	13.2	9.4
Moderate bull (2025)	13.2	7.6	5.6
Bear (2022)	-27.0	-8.2	-18.8

Source: Calculated from the fund-year panel.

The two-way ANOVA provides an important insight: the interaction between investment approach and market condition is statistically significant ($F = 11.6$, $p < 0.001$). In other words, neither approach has the same advantage in every type of market. During strong bull-market years, ML-based funds returned around 9.4 percentage points more than traditional funds, and their advantage was about 5.6 points in 2025. However, in 2022 they lost approximately 18.8 percentage points more. These contrasting outcomes help explain why the overall average-return comparison was not statistically significant. This pattern differs from Babiak and Barunik's (2026) finding that deep-learning gains were strongest during recessions. In this sample, the ML-based funds behaved more like higher-beta equity portfolios. Since only one bear-market year is included, the result should be interpreted cautiously.

Multiple linear regression

Table 10: Regression of Annual Return on Approach, RME and Market Condition (H01)

Variable	Coefficient	Std. error	t	p-value	VIF
Intercept	-8.85	6.78	-1.31	0.203	n/a
Approach (ML=1)	11.11	12.91	0.86	0.397	18.0
Risk management efficiency	4.63	7.14	0.65	0.523	18.0
Market condition (S&P 500 return)	0.79	0.09	9.03	<0.001	1.0
R-squared / Adjusted R-squared	0.761	0.734			
F-statistic (p-value)	27.66	<0.001			
Durbin-Watson	2.42				

Source: Calculated from the fund-year panel ($n = 30$).

The regression model is statistically significant overall ($F = 27.7$, $p < 0.001$) and accounts for approximately 76% of the variation in annual returns. On this basis, the study rejects H01. However, market conditions are the only individually significant predictor. A one-percentage-point increase in the S&P 500 return is associated with an increase of about 0.79 percentage points in fund returns ($p < 0.001$). After accounting for market conditions, neither the approach indicator ($p = 0.397$) nor RME ($p = 0.523$) is statistically significant. The variance inflation factors of around 18 indicate substantial overlap between approach and RME: all three ML-based funds also have the weakest risk-efficiency scores. As a result, the independent contribution of each variable cannot be estimated reliably. The Durbin–Watson statistic of 2.42 does not suggest serious autocorrelation.

Effect of risk management efficiency and mediation

Table 11: Effect of Risk Management Efficiency on Performance (H04)

Model	Coefficient on RME	Std. error	t	p-value	R-squared
Annual return on RME (fund-years, $n = 30$)	-1.35	3.31	-0.41	0.688	0.006
Mean return on RME (funds, $n = 6$)	-1.35	0.95	-1.42	0.228	0.336
Sharpe ratio on RME (funds, $n = 6$)	0.038	0.048	0.80	0.469	0.138

Source: Calculated from Tables 3 and 4.

Table 12: Mediation Analysis (Approach to RME to Return)

Path	Estimate	p-value
------	----------	---------

Path	Estimate	p-value
a: Approach (ML = 1) to RME	-1.76	0.001 (fund level)
c: Approach to Return (total effect)	2.98	0.622
b: RME to Return, controlling for approach	4.63	0.748
c': Approach to Return (direct effect)	11.11	0.670
Indirect effect (a x b); Sobel z	-8.13; -0.32	0.746

Source: Calculated from the fund-year panel. Path a is tested on the six funds because RME is the same in every year for a fund.

Across the three tests, RME did not have a statistically significant relationship with performance, so H04 is not rejected. Its estimated relationship with annual returns was negative but insignificant ($p = 0.688$). Across the six funds, stronger risk control tended to accompany lower average returns (rank correlation = -0.54), which may reflect the usual trade-off between risk and return. The relationship between RME and the Sharpe ratio was slightly positive but also insignificant ($p = 0.469$). The mediation test found that the optimisation approach was associated with RME, but RME did not significantly transmit an effect to returns. The indirect effect was not significant (Sobel $z = -0.32$, $p = 0.746$), and the total effect of approach on returns was also insignificant. Therefore, the proposed mediating role is not supported by this dataset. The small sample, however, means that a genuine relationship may have gone undetected.

Findings and Discussion

- H01 (rejected): The regression of annual return on approach, risk management efficiency and market condition is significant (R-squared = 0.76), but only the market condition is individually significant.
- H02 (rejected, with caution): Risk management efficiency is significantly higher for the traditional minimum-variance and low-volatility funds ($F = 67.9$, $p = 0.001$). They had lower volatility, smaller losses in 2022 and lower betas.
- H03 (not rejected on a pooled basis): Average annual return, active return, Sharpe ratio and Sortino ratio do not differ significantly between ML-based and traditional funds. The relationship does depend on the market, since ML-based funds earned more in rising markets and lost far more in the 2022 decline (interaction $p < 0.001$).
- H04 (not rejected): Risk management efficiency has no significant effect on performance, and no mediation was found.
- Neither group matched the S&P 500: both underperformed it on average by 5 to 8 percentage points a year.

Overall, the findings sit between the contrasting conclusions reported in earlier studies. Studies such as Gu et al. (2020), Ngo et al. (2023) and Wang et al. (2026) report that ML improves performance. This study finds that three commercial AI-driven funds earned a higher average return than traditional funds but not a better risk-adjusted return, and that they did not manage risk better. This supports the caution of DeMiguel et al. (2009) and Michaud (1989) that sophisticated models do not necessarily beat simpler strategies, and it is consistent with the view of López de Prado et al. (2025) that implementation challenges remain. The funds examined are marketed products, which charge fees and may use AI only for part of the selection process, so the results do not show that ML as a technique is inferior. The study also shows that a traditional minimum-variance approach delivers what it promises, namely lower risk.

Limitations of the Study

- The sample has six funds and five annual observations each, so the fund-level tests have low power and the fund-year observations are not independent. P-values should be treated as indicative.
- Risk measures are based on annual returns and not on daily or monthly data, so true volatility and maximum drawdown within a year could not be calculated.
- The period contains only one bear market year (2022), so conclusions about downside behaviour rest on one episode.
- The ML-based funds are commercial products with fees, and AMOM also carries a momentum tilt. The traditional funds use index rules and include a global fund (ACWV) that is not directly comparable with the U.S. benchmark.
- Risk management efficiency is constant within a fund, and it is strongly collinear with the approach dummy (variance inflation factor of about 18), so their separate effects cannot be separated.
- Returns were taken from public websites on a market-price basis, and small differences from net asset value returns are possible.

Suggestions and Recommendations

- Investors should judge AI-driven products by their risk-adjusted results through full market cycles, including a downturn, and not by return in rising markets alone.
- Investors who prefer stability may find that minimum-variance and low-volatility funds provide more reliable protection in falling markets, while accepting lower returns when markets rise.
- Fund providers should disclose how AI is used in portfolio selection and how risk is controlled, so that products can be compared with traditional alternatives.

- Portfolio managers may consider using ML to improve the inputs of a traditional optimiser or to add explicit risk constraints, which the literature suggests can combine the strengths of both approaches (Wang et al., 2020; Ma et al., 2021).

Future studies could use daily or monthly data and a longer period, add more funds and Indian equity funds, and include transaction costs. They could also construct ML and traditional portfolios on the same asset universe, so that the optimisation approach is the only difference.

Conclusion

This study assessed six ETFs over the five-year period from 2021 to 2025 to compare AI/ML-based portfolio approaches with traditional minimum-variance and low-volatility strategies. The traditional funds showed stronger risk control, reflected in lower volatility, smaller losses during 2022 and lower beta values. Although ML-based funds performed better in rising markets, their larger losses during the 2022 downturn offset those gains, leaving no statistically significant difference in average returns between the groups. Risk-management efficiency did not significantly predict performance, and the proposed mediation relationship was not supported. Market conditions were the strongest predictor of returns in the regression model. These conclusions are limited by the small sample and the inclusion of only one bear-market year. Future research using a longer period, more funds and more frequent return data could provide a clearer picture of when AI-based approaches offer an advantage over traditional optimisation.

REFERENCES :

- Agal, S., Raulji, K., & Odedra, N. D. (2025). A machine learning approach to risk based asset allocation in portfolio optimization. *Scientific Reports*.
- Babiak, M., & Baruník, J. (2026). Deep learning, predictability, and optimal portfolio returns. *Journal of Empirical Finance*.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.
- Betancourt, C., & Chen, W.-H. (2021). Deep reinforcement learning for portfolio management of markets with a dynamic number of assets. *Expert Systems with Applications*.
- Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*.
- Board of Governors of the Federal Reserve System. (2026). 3-month Treasury bill secondary market rate, discount basis [RIFSGFSM03NA]. FRED, Federal Reserve Bank of St. Louis.
- DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*.
- Freyberger, J., Neuhiel, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*.
- Gollart, M., & Okhrin, Y. (2025). A reinforcement learning approach to dynamic portfolio optimization. *Annals of Operations Research*.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*.
- Ha, Y. S., Kang, J., Kim, J., Chun, D., et al. (2026). Machine learning-based portfolio optimization: Comparative analysis with the All-Weather portfolio strategy. *Financial Innovation*.
- Heaton, J. B., Polson, N. G., & Witte, J. H. (2016). Deep learning in finance. *arXiv*.
- Jang, J., & Seong, N. (2023). Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory. *Expert Systems with Applications*.
- Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *arXiv*.
- Kelly, B., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*.
- Kozak, S., Nagel, S., & Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*.
- Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*.
- López de Prado, M., Simonian, J., Fabozzi, F. A., & Fabozzi, F. J. (2025). Enhancing Markowitz's portfolio selection paradigm with machine learning. *Annals of Operations Research*.
- Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*.
- Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*.
- Ngo, V. M., Nguyen, H. H., & Nguyen, P. V. (2023). Does reinforcement learning outperform deep learning and traditional portfolio optimization models in frontier and developed financial markets? *Research in International Business and Finance*.
- Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1), 119-138.
- Simon, F., Weibels, S., & Zimmermann, T. (2026). Deep parametric portfolio policies. *Management Science*.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290-312.
- Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. *The Journal of Investing*, 3(3), 59-64.
- Wang, W., Li, W., Zhang, N., & Liu, K. (2020). Portfolio formation with preselection using deep learning from long-term financial data. *Expert Systems with Applications*.
- Wang, Y., Gao, H., Harvey, C. R., Liu, Y., & Tao, X. (2026). Machine learning meets Markowitz (NBER Working Paper No. 34861). National Bureau of Economic Research.
- Zhang, Z., Zohren, S., & Roberts, S. (2020). Deep learning for portfolio optimization. *arXiv*.