

When Strategies Work Too Well: Volatility-Based Investing, Overfitting and the Limits of Empirical Finance

Francisco Bogado Valls*, Elmar Steurer*^{ID}

Faculty of Business and Management, Hochschule Neu-Ulm, Neu-Ulm, Germany

Email: *francisco.bogado.valls@gmail.com, *Elmar.Steurer@hnu.de

How to cite this paper: Valls, F. B., & Steurer, E. (2026). When Strategies Work Too Well: Volatility-Based Investing, Overfitting and the Limits of Empirical Finance. *Journal of Financial Risk Management*, 15, 189-211.

<https://doi.org/10.4236/jfrm.2026.153012>

Received: April 20, 2026

Accepted: July 10, 2026

Published: July 13, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0). <http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

This study examines the performance and robustness of an index volatility portfolio allocation strategy applied to the S&P 500, DAX, and Nikkei 225 indices for the period 1996-2025. The volatility portfolio allocation strategy is based on a dynamic market allocation that uses discrete volatility regimes estimated according to historical data, so that the portfolio enters or exits the market entirely depending on the level of observed volatility. Empirical results show that volatility portfolio allocation rules can achieve substantial improvements in cumulative returns and risk-adjusted performance compared to a passive Buy-and-Hold strategy in various financial markets. The strategy often achieves superior performance despite being invested for a smaller percentage of time. However, subsequent analysis reveals significant methodological limitations that affect the interpretation of the results. The strategy performance is highly sensitive to parameter selection, exhibiting instability between sub-periods and a high dependence on model specification. The study shows the existence of various methodological problems, such as look-ahead bias in the contemporaneous (t-0) specification, extensive in-sample parameter optimization, and the absence of out-of-sample validation. All of this creates concerns related to data mining, backtest overfitting, and the economic robustness of the strategy. Overall, the results suggest that the improvements observed in performance do not constitute strong evidence of a stable relationship between volatility and future returns. The results highlight the importance of rigorous methodology and the evaluation of robustness in empirical financial research. The study contributes to the literature by exposing the risks involved in the selection of data-driven models.

Keywords

Volatility Timing, Portfolio Allocation, Backtest Overfitting, Data Mining,

1. Introduction

Market volatility plays a leading role in the theory of asset pricing, risk management, and portfolio allocation. Financial literature shows that the volatility of returns follows a stochastic process in which volatility is dependent, shows persistence and clustering, meaning that periods of high volatility are followed by periods of high volatility as well (Cont, 2001). These properties are of substantial importance in financial theory and have inspired the development of models that represent volatility as a dynamic, perhaps predictable process (Andersen, 2001).

In this sense, multiple studies have focused on whether the temporal variation in volatility levels can be applied to improve investment decisions. Certain strategies, commonly referred to as volatility timing strategies, have been developed that dynamically adjust market exposure, using historical or expected measures of volatility, with the purpose of finding a better risk-adjusted return in a portfolio (Moreira & Muir, 2017). It has also been suggested that volatility contains valuable information in the dynamic allocation of assets and that it can have value in risk management (Fleming et al., 2003). On the other hand, empirical evidence on the relationship between volatility and future returns is ambiguous. While some studies suggest that dynamically adjusting risk exposure provides benefits, others find that the relationship between conditional volatility and expected returns is weak or unstable (Baillie & DeGennaro, 1990). This lack of consensus has raised greater concerns about the robustness of empirical results in finance and the danger of using historical data as a method for designing investment strategies.

Empirical research in financial markets finds a series of important methodological problems related to the availability of large data sets and the possibility of studying, evaluating, and even comparing many models and parameters. Intensive data exploration may lead to the conclusion that there are statistical patterns that are inauthentic and do not reveal authentic economic relationships. Specifically, the phenomenon designated as backtest overfitting refers to the tendency of investment strategies optimized from historical data to have premium performance results due to overfitting certain sample characteristics (Bailey et al., 2015). Also, selecting strategies based on multiple tests can lead to selection biases and cause historical performance to be overestimated (Bailey & de Prado, 2014).

This work aims to contribute to this issue by providing an empirical analysis of the development and evaluation process of a volatility regime-based portfolio allocation strategy. Using daily data from the S&P 500, DAX, and Nikkei 225 indices from 1996 to 2025, a discrete investment rule is first presented that adjusts market exposure according to certain ranges of historical volatility. The primary objective is to determine whether this type of allocation rule can improve portfolio performance in terms of cumulative returns and adjusted performance. However, dur-

ing the investigation, certain methodological problems that have a significant impact on the results obtained begin to be detected. Specifically, look-ahead bias in the contemporaneous (t-0) specification, intensive parameter optimization on the same sample of data obtained, and the lack of out-of-sample validation are detected. Therefore, the contemporaneous (t-0) specification should be viewed more as a diagnostic benchmark than as a strategy that can be implemented for trading. These factors can result in a series of outcomes that, although external, do not reflect real economic regularities. The contribution of this work is twofold: on the one hand, it documents the fragility of volatility-based investment strategies when subjected to realistic methodological constraints; on the other, it provides evidence on the economic role of volatility as a measure of market uncertainty and not as a predictor of future returns.

The study's findings highlight the importance of methodological rigor in empirical financial research and demonstrate that quantitative strategies are extremely sensitive to biases in the modeling decision and parameter selection. In addition to the methodological contribution, this work also provides evidence on the economic interpretation of market volatility. Empirical findings suggest that volatility is primarily explained by market uncertainty and changing market conditions, rather than by systematic variation in expected returns. The lack of stable predictive patterns between volatility and future returns reinforces the view that volatility behaves as a market state variable, which reflects the level of uncertainty in the market, and not as a reliable predictor of risk premiums.

2. Theoretical Background

2.1. Financial Returns and Volatility

The study of financial returns represents the central axis of the empirical study of asset pricing and portfolio management research. Literature has shown that financial returns have specific statistical properties (non-normal distribution, heavy tails, heteroscedasticity, and temporal persistence of volatility) which, added to the phenomenon of clustered volatility, refer precisely to the fact that periods of high volatility are often followed by another period of high volatility (Cont, 2001). The time variation of volatility becomes relevant to financial theory when it is argued that market risk is not constant over time and can impact investment decisions or asset valuation (Andersen, 2001).

2.2. Measurement of Volatility

The empirical measure of volatility represents one of the central aspects of financial research. Literature covers a range of approaches to estimate the volatility of returns based on different statistical assumptions and the information used. The conditional heteroskedasticity model, present in the ARCH and GARCH models, proposes the variance of returns as a function of past information, allowing the incorporation of the temporal persistence of volatility (Engle, 1982; Bollerslev, 1986). These types of models are applied to model the dynamics of financial risk.

On the other hand, there are also non-parametric measures of volatility based on historical data. Historical volatility can be estimated through the standard deviation of returns in moving time windows, reflecting the market's most recent variability. Similarly, more efficient estimators have emerged from additional information about prices. Parkinson's estimator uses high and low prices to enhance the estimation of variance, (Parkinson, 1980). while the Garman-Klass estimator includes information on opening, closing, extreme prices, and the maximum and minimum prices of the period (Garman & Klass, 1980). Another widely used way to estimate volatility and risk is the use of exponential smoothing methods, such as the EWMA model (Longerstaey & Spencer, 1996). The fact that there are different estimation methods is a clear indication of the complexity of risk measurement.

2.3. Risk-Return Relationship

The risk-return relationship is one of the fundamental pillars of financial theory. Classic asset pricing models assume that investors want to receive a higher expected rate of return if they assume a higher level of risk, such that there is a positive relationship between volatility and expected return. However, empirical findings on this relationship are inconsistent. Some studies provide little evidence of a positive relationship between conditional volatility and future returns (Baillie & DeGennaro, 1990). Other works show that the risk-return relationship can change over time and may depend on the econometric specification (Kenneth R. French et al., 1987). Consequently, this lack of correlation has led to further research on the predictive power of volatility and its role in investment decisions and portfolio composition.

2.4. Volatility Timing and Dynamic Portfolio Allocation

The timing of volatility has given rise to investment strategies that create the allocation of capital whose long positions on volatility are based on historical or expected volatility. Literature shows that dynamic risk exposure management results in the possibility of a higher outcome than the portfolio's average composition if it occurs under the conditions that literature requires (Moreira & Muir, 2017). On the other hand, it has been shown that volatility can have economic value when making asset allocation and risk management decisions (Fleming et al., 2003). However, recent studies have shown that the profitability of volatility decisions will depend significantly on the investment horizon, the model specification, and the sample period (Moreira & Muir, 2019). This means that the robustness of the profitability of volatility-based investment strategies becomes an open empirical question.

2.5. Data Mining and Backtest Overfitting in Empirical Finance

Empirical research in finance also has methodological risks associated with the intensive exploitation of historical data. Repeatedly searching for models and pa-

rameters can lead to the successful detection of a false statistical relationship that does not play a role in real economic mechanisms. The phenomenon of backtest overfitting corresponds to the inflated profitability of investment strategies chosen from historical data due to overfitting to the specificity of the sample (Bailey et al., 2015). Furthermore, research with multiple specifications can also lead to selection bias and overestimation of historical profitability (Bailey & de Prado, 2014). From a statistical standpoint, the validity of empirical results depends on the robustness of the model to specification errors, outliers, and changes in the data (Maronna et al., 2006). Considering methodological arguments is essential for evaluating the validity of quantitative strategies and is an important part of the empirical reflection developed throughout this work.

3. Data and Methodology

3.1. Data

The empirical analysis in this paper is based on daily prices of the S&P 500, DAX, and Nikkei 225 stock index series for the period between January 1996 and December 2025. These series correspond to developed markets in the United States (S&P 500), Europe (DAX), and Asia (Nikkei 225), allowing for a true assessment of the robustness of results in different financial environments.

The data were extracted from the Stooq.com database using the index tickers ^SPX, ^DAX, and ^NKX. The series correspond to price indices expressed in their respective local currencies (USD, EUR, and JPY). The data were used as provided by Stooq, without additional adjustments for corporate actions or currency conversion.

As each index is evaluated individually, no synchronization procedure was applied to align trading days across markets. Therefore, the returns and volatility measures for each index were computed using the appropriate trading calendar for that index.

Also, the use of daily data allows for the proper capture of the short-term dynamics of financial markets and makes it possible to analyze the relationship between historical volatility and future returns with greater statistical precision (Cont, 2001).

3.2. Return Calculation

The econometric study of prices and returns establishes the theoretical framework that allows the analysis of the evolution of financial markets and the possible predictability of their fundamental variables. Market returns are calculated as simple returns based on daily closing prices, with the following mathematical expression: the return of the index in period t defined as

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

where P_t represents the closing price of the index in period t . Since it is the

standard definition in which financial econometrics considers variations in asset prices, it allows the empirical analysis of the dynamics of risk and return (Campbell et al., 1998).

3.3. Volatility Estimation

The volatility was estimated using historical volatility calculated in rolling windows. The rolling volatility is given by the standard deviation of returns in a 100-day window, annualized by the corresponding scale factor:

$$\hat{\sigma}_t = \sqrt{250 \cdot \frac{1}{100} \sum_{i=0}^{99} r_{t-i}^2}$$

A measure considered a non-parametric approximation of market risk based on the variability of the most recent returns (Longerstaey & Spencer, 1996). The choice of historical volatility as a risk estimator is motivated by empirical evidence of the existence of temporal persistence for financial volatility, as well as its widespread application in risk management and portfolio allocation (Andersen, 2001).

3.4. Volatility-Based Investment Strategy

This paper presents a dynamic portfolio allocation strategy, based on discrete regimes of asset volatility. The strategy adjusts market exposure according to the level of volatility estimated from historical return volatility, so that it allows for total market entry or exit depending on the estimated volatility regime. The portfolio return is defined as

$$r_{p,t} = w_t \cdot r_{m,t}$$

where $r_{m,t}$ is the return of the index and w_t is the market exposure.

Portfolio returns in this model are assumed to equal zero when the strategy has exited the market ($w_t = 0$). This assumption allows the analysis to isolate the effect of volatility-based timing decisions without introducing assumptions regarding risk-free returns.

Exposure to the market is reduced by applying discrete rules based on volatility ranges:

$$w_t = \begin{cases} 1 & \text{if } \hat{\sigma}_t \in [\sigma_L, \sigma_M] \\ 0 & \text{if } \hat{\sigma}_t \in (\sigma_M, \sigma_H) \\ 1 & \text{if } \hat{\sigma}_t > \sigma_H \end{cases}$$

Volatility thresholds are selected empirically through the analysis of historical volatility buckets constructed in one-percentage-point intervals. The strategy excludes continuous volatility ranges associated with weaker historical performance. The proposal is linked to “volatility timing” strategies, so that risk exposure is dynamically adjusted according to market volatility measures (Moreira & Muir, 2017).

Two specifications are considered. In the contemporaneous (t-0) specification, exposure is determined using volatility estimated in the same period, which intro-

duces look-ahead bias. In the lagged (t+1) specification, volatility estimated at time t determines market exposure in period t+1, making the strategy implementable in real time.

3.5. Parameter Selection and Model Optimization

The model parameters are selected by evaluating the historical performance of the strategy for different volatility ranges constructed in one-percentage-point intervals. The parameter-selection process involved an iterative exploratory analysis in which alternative volatility thresholds were evaluated manually according to cumulative portfolio returns across indices and model specifications. This procedure involves exploring multiple model configurations with the goal of maximizing historical performance. Intensive parameter optimization on the same data sample can generate optimistic estimates of historical performance due to overfitting to specific characteristics of the data, a phenomenon known as backtest overfitting (Bailey et al., 2015). Additionally, evaluating multiple specifications can produce selection biases and non-replicable results (Bailey & de Prado, 2014).

3.6. Predictive Regression Framework

To analyze whether historical volatility is a good predictor of future returns, we ran a series of linear predictive regressions for each of the indexes analyzed in our study. We investigated the relationship between realized historical volatility and future index returns based on two time horizons (one day and five days). The model is specified as follows:

$$r_{t+h} = \alpha + \beta \sigma_t + \varepsilon_t$$

where r_{t+h} represents the future index return at horizon h , σ_t denotes historical volatility at time t , and ε_t is the error term. The coefficient β measures the strength of the relationship between volatility and future returns. If there is a statistical significance on the coefficients, then volatility will have some ability to predict future returns. The volatility measure corresponds to the same 100-day rolling annualized volatility estimator used in the portfolio allocation strategy. The models are estimated using ordinary least squares (OLS) for two different forecast horizons: $h = 1$ (next-day return) & $h = 5$ (five-day return). Non-overlapping observations were used in order to avoid creating mechanical autocorrelation from overlapping return windows in the five-day specification. Given the potential presence of heteroskedasticity in financial return series, inference should be interpreted with caution.

Data from each day for the years 1996 through to 2025 are used to determine daily returns for the S&P 500, DAX and Nikkei 225. The relationship between conditional volatility and future returns has been extensively studied in financial literature, although in general, empirical evidence on the predictive power of volatility is scarce and dependent on the model used (Baillie & DeGennaro, 1990).

3.7. Methodological Considerations

Empirical analysis should consider potential methodological problems that may be associated with the use of historical data, including the look-ahead bias resulting from the use of contemporaneous information, the practice of in-sample optimization, and the lack of out-of-sample validation, which can lead to spurious empirical results and overestimation of the behavior of quantitative strategies (Bailey et al., 2015). The criticism of the methodological problems discussed here is a central part of this work.

4. Empirical Results

4.1. Overview of Empirical Performance

The empirical results of the dynamic volatility-based strategy are presented for the S&P 500, DAX, and Nikkei 225 over the period 1996–2025. The results indicate that the VM strategy outperformed B&H in terms of cumulative returns, annualized return, portfolio volatility, and market exposure. Performance differences were substantial across both model specifications and markets (see Appendix).

4.2. Aggregate Performance Results

Table 1. Aggregate performance (1996–2025).

Panel A: S&P 500			
Model	Buy-and-Hold	VM (t-0)	VM (t+1)
Volatility Regime Rule (%)	—	0 - 15 (In)	0 - 15 (In)
		16 - 23 (Out)	16 - 17 (Out)
		24+ (In)	18+ (In)
Cumulative Return (%)	1011.41	1583.37	1697.99
Annual Return (%)	8.36	9.87	10.11
Portfolio Volatility (%)	19.03	15.34	18.07
Days in Market	7550	5204	6858
Days out of the Market	0	2346	692
Time Invested (%)	100	68.93	90.83
Panel B: DAX			
Model	Buy-and-Hold	VM (t-0)	VM (t+1)
Volatility Regime Rule (%)	—	0 - 17 (In)	0 - 22 (In)
		18 - 29 (Out)	23 - 30 (Out)
		30+ (In)	31+ (In)
Cumulative Return (%)	986.58	1361.52	1506.83
Annual Return (%)	8.28	9.35	9.7
Portfolio Volatility (%)	22.38	16.57	19.44
Days in Market	7607	4516	6224
Days out of Market	0	3091	1383
Time Invested (%)	100	59.37	81.82

Continued

Panel C: Nikkei 225			
Model	Buy-and-Hold	VM (t-0)	VM (t+1)
Volatility Regime Rule (%)	—	0 - 18 (In)	0 - 16 (In)
		19 - 29 (Out)	17 - 20 (Out)
		30+ (In)	21+ (In)
Cumulative Return (%)	153.37	420.79	387.65
Annual Return (%)	3.15	5.65	5.42
Portfolio Volatility (%)	23.1	15.65	20.53
Days in Market	7357	3559	5465
Days out of the Market	0	3798	1892
Time Invested (%)	100	48.38	74.28

Interpretation (Table 1)

In summary, this study shows that a volatility-managed (VM) strategy outperforms a passive strategy with respect to cumulative returns across each of the three stock indices examined (S&P 500, DAX, and Nikkei). For example, cumulative returns for the S&P 500 increased from 1.011% under the Buy-and-Hold (B&H) strategy to 1.583% and 1.698% under the t-0 and t+1 VM specifications. In addition to generating higher cumulative returns than the passive benchmark, the VM strategy also generally produced superior risk-adjusted performance.

4.3. Market Exposure Analysis**Table 2.** Market exposure.

Index	Model	Days in the Market	Days out of the Market	Time Invested (%)
S&P 500	VM (t-0)	5204	2346	68.93
S&P 500	VM (t+1)	6858	692	90.83
DAX	VM (t-0)	4516	3091	59.37
DAX	VM (t+1)	6224	1383	81.82
Nikkei 225	VM (t-0)	3559	3798	48.38
Nikkei 225	VM (t+1)	5465	1892	74.28

Interpretation (Table 2)

The volatility-managed model exhibits lower market exposure than the passive benchmark. The fact that the strategy outperformed Buy-and-Hold may be partially due to excluding intermediate volatility regimes. Volatility thresholds also exhibit a great deal of variability across market exposures within each model and between markets, indicating strong sensitivity to volatility-threshold selection.

4.4. Subperiod Analysis

The temporal stability of the model is evaluated based on strategy performance across three ten-year subperiods (as shown in Table 3): 1996-2005, 2006-2015,

and 2016–2025. The subperiod analysis is used to determine whether the full-sample results reflect stable structural relationships or depend on specific market conditions. Temporal stability therefore represents an important criterion in the evaluation of the robustness of investment strategies.

Table 3. Subperiod performance.

Panel A: S&P 500				
Model	Period	Cumulative Return (%)	Annual Return (%)	Portfolio Volatility (%)
Buy-and-Hold	1996–2005	102.67	7.32	18.25
VM (t-0)	1996–2005	135.64	8.95	12.21
VM (t+1)	1996–2005	199.24	11.58	16.75
Buy-and-Hold	2006–2015	63.74	5.05	20.67
VM (t-0)	2006–2015	89.94	6.63	17.80
VM (t+1)	2006–2015	68.80	5.38	20.10
Buy-and-Hold	2016–2025	234.92	12.85	18.06
VM (t-0)	2016–2025	276.11	14.16	15.50
VM (t+1)	2016–2025	255.94	13.54	17.19
Panel B: DAX				
Model	Period	Cumulative Return (%)	Annual Return (%)	Portfolio Volatility (%)
Buy-and-Hold	1996–2005	139.95	9.15	25.24
VM (t-0)	1996–2005	180.91	10.88	18.65
VM (t+1)	1996–2005	193.25	11.36	21.20
Buy-and-Hold	2006–2015	98.64	7.10	22.89
VM (t-0)	2006–2015	124.92	8.44	17.01
VM (t+1)	2006–2015	118.45	8.13	20.37
Buy-and-Hold	2016–2025	127.97	8.59	18.49
VM (t-0)	2016–2025	131.32	8.75	13.65
VM (t+1)	2016–2025	150.84	9.63	16.43
Panel C: Nikkei 225				
Model	Period	Cumulative Return (%)	Annual Return (%)	Portfolio Volatility (%)
Buy-and-Hold	1996–2005	−18.91	−2.07	22.85
VM (t-0)	1996–2005	23.93	2.17	12.21
VM (t+1)	1996–2005	9.19	0.88	20.97
Buy-and-Hold	2006–2015	18.14	1.68	25.37
VM (t-0)	2006–2015	23.17	2.11	18.62
VM (t+1)	2006–2015	20.25	1.86	23.30
Buy-and-Hold	2016–2025	164.48	10.21	20.83
VM (t-0)	2016–2025	241.17	13.06	15.45
VM (t+1)	2016–2025	271.40	14.02	16.77

Interpretation:

Subperiod data shows that the relative performance of the volatility-based strategy can differ over time. The strategy shows an overall tendency to outperform the Buy-and-Hold benchmark, but both the magnitude of outperformance and which specification performs best depend upon the specific time period examined. This evidence suggests that the parameters of the model are unstable and that its performance depends strongly on current market conditions.

4.5. Sensitivity to Volatility Thresholds

The results indicate that the volatility intervals that determine market exposure are different depending on the index but also depending on the model specifications. For example, ranges that exclude the market are between 16% and 23% for the S&P 500; 18% and 29% for the DAX; and 19% and 29% for the Nikkei under the contemporaneous (t-0) specification. The high dependence of performance on the choice of thresholds also highlights the high sensitivity of the model to parameter specification, as relatively minor changes in volatility intervals result in very large changes in portfolio performance. This dependence is consistent with previous evidence of the fragility of strategies that have been optimized through the exploration of historical data (Bailey et al., 2015).

4.6. Statistical Tests

Statistical tests were used to evaluate whether the returns associated with the Volatility Managed (VM) Strategy differed significantly from those generated by the passive Buy-and-Hold (BH) Benchmark.

Return differentials:

$$dt = r_{VM,t} - r_{BH,t}$$

between the VM strategy and the BH benchmark were assessed for statistical significance.

Table 4. HAC/Newey-west test results (Lag = 5): volatility managed strategy vs buy-and-hold.

Index	Test	VM (t-0)	VM (t+1)
S&P 500	HAC t-statistic	0.3901	1.3578
S&P 500	HAC <i>p</i> -value	0.6965	0.1746
DAX	HAC t-statistic	-0.0598	0.3482
DAX	HAC <i>p</i> -value	0.9523	0.7277
Nikkei 225	HAC t-statistic	0.3379	0.8458
Nikkei 225	HAC <i>p</i> -value	0.7354	0.3977

The statistical significance of the return differentials between the two strategies was estimated using heteroskedasticity and autocorrelation consistent (HAC) standard errors following the Newey and West methodology (as shown in **Table**

4). A total of five lagged terms were utilized for estimation. The Newey-West methodology is particularly well suited for financial time-series analysis because it accounts for both heteroskedasticity and short-run serial correlation.

Interpretation:

The HAC/Newey-West adjusted results provide limited statistical evidence that the average return differentials between the volatility-managed strategy and the passive benchmark are significantly different from zero. Although the volatility-managed strategy produced substantial differences in cumulative portfolio performance, the statistical evidence for persistent average daily excess returns is weaker after correcting for heteroskedasticity and autocorrelation. These findings are consistent with the broader methodological conclusions of this study, namely that strong historical backtest performance does not necessarily imply the existence of a stable or economically robust relationship between volatility and future returns.

4.7. Predictive Regression Results

Historical volatility does not have predictive value for the next day's return across all three stock indexes used in the study. The regression of DAX returns on historical volatility has a small ($\beta = 0.0011$) and statistically insignificant (p -value = 0.545) coefficient, and its R-squared statistic is approximately zero. The S&P 500 index is similar, the coefficient of historical volatility is also statistically insignificant ($\beta = 0.0015$, p -value = 0.336), and it has virtually no ability to explain variance in daily returns and its R-squared statistic is also approximately zero. The Nikkei 225 index also exhibited no statistically significant relationship between returns and volatility ($\beta = 0.0014$, p -value = 0.541), and its R-squared was also essentially zero. In each case, the data suggests that past volatility explains none of the variation in next-day returns.

On the five-day horizon, the results do not show much evidence of any strong or consistent relationship between volatility and subsequent returns. The S&P 500 had a statistically significant coefficient on volatility ($\beta = 0.008$, p -value = 0.015), however the R-squared was only 0.0008, which indicates very little explanatory power. There was no statistically significant relationship between volatility and Nikkei 225 returns ($\beta = 0.0049$, p -value = 0.296). DAX had also no statistically significant coefficient on volatility ($\beta = 0.005$, p -value = 0.16) and the R-squared was only 0.0003.

Ultimately, the results provide little evidence to support the idea that historical volatility is a reliable predictor of future returns. There appears to be little or no predictive value at short horizons, and at longer horizons there were some statistically significant relationships but had low explanatory power. These results tend to support the conclusion that volatility primarily reflects current levels of uncertainty about the market rather than systematic differences in expected returns and future returns.

4.8. Economic Interpretation of Results

The prevailing empirical results suggest that the volatility-based strategy can lead

to significant improvements in portfolio performance relative to the passive strategy in different market environments, by comparing the performance of the two strategies in different specifications and time intervals. One result worth noting is that the superior performance occurred even with less market exposure. This is because the strategy avoids certain volatility regimes and is out of the market for a relatively high percentage of the time, suggesting that the exclusion of selected market states may be part of the reason for the superior performance. The analysis also shows a strong dependence on the choice of the parameters and specifications for the model. The optimal volatility intervals vary depending on the indices, specifications, and dates. Additionally, slight differences in thresholds will result in substantial variations in the results. This set of positive returns and high sensitivity to parameters does not necessarily show a structural economic relationship between volatility and future returns but may depend on the characteristics of our sample.

4.9. Summary of Empirical Results

The empirical results of this analysis yield three main findings. First, the volatility-based strategy significantly improves portfolio returns relative to the passive strategy in the analyzed markets. Second, the model's performance is highly sensitive to the choice of volatility thresholds and is sensitive to variations in subperiods. Third, the strategy exhibits improvements in performance even with reduced market exposure.

5. Methodological Issues and Model Robustness

5.1. Introduction

The empirical results show that the volatility-related strategy can lead to considerable improvements in portfolio performance but also show that the strategy is very sensitive to the parameter and the specification of the procedure used. This chapter highlights the main methodological problems related to the development of the strategy, look-ahead bias in the development of the investment signal, the process of parameter selection through intensive data exploration, as well as the lack of validation outside the sample.

5.2. Look-Ahead Bias

One of the methodological problems is directly related to the use of contemporary information in estimating the volatility used for investment decision-making. This problem is known as look-ahead bias, since the model is using information that would not be available in real time, which artificially inflates the strategy's estimates (Bailey et al., 2015). In the first specification (t-0), historical volatility was estimated by taking the returns r_t . Subsequently, the volatility estimate was used to define market exposure in the same period. Formally:

$$w_t = f(\hat{\sigma}_t)$$

$$r_{p,t} = w_t r_{m,t}$$

This practice is incompatible with the requirement of temporal causality, since the investor cannot know the return for the period when they make the decision (Campbell et al., 1998). To address the look-ahead bias, an alternative is specified in which the historical volatility estimated in period t is what determines exposure in period $(t + 1)$.

5.3. Data Mining and Parameter Selection

A second methodological issue has to do with the process of selecting the model parameters. The volatility intervals that determine market exposure were selected through an iterative empirical exploration of multiple configurations and analyzing the strategy past performance under different volatility ranges. This methodology involves data mining, in which the parameters are estimated ex post based on their performance in the sample analyzed. Literature has documented that such data exploration increases the likelihood of finding false statistical patterns that do not translate to genuine economic relationships (Bailey et al., 2015). Similarly, evaluating multiple specifications without correcting for multiple testing can bias the selection and overrepresent the statistical significance of the results (Bailey & de Prado, 2014). Empirical analysis leads to the conclusion that the volatility intervals that produce improvements in performance are not the same across indices and across model specifications, which indicates no structurally stable parameters.

5.4. Parameter Instability and Model Fragility

A result of the analysis is the model's high sensitivity to the choice of volatility thresholds. A small variation in the decision intervals yields substantial differences in portfolio profitability, suggesting that the model is fragile. Additionally, there are parameters that will generate performance in some periods, but not in others. This temporal instability indicates that the model does not capture a structural relationship between volatility and future returns but rather depends on specific sample characteristics. The high sensitivity to model specification is consistent with previous evidence on the fragility of strategies derived from intensive historical data searches (Bailey et al., 2015).

5.5. Absence of Out-of-Sample Validation

The process of estimating and evaluating the model was performed solely on the same sample without distinction between training and validation periods. The lack of out-of-sample validation prevents us from evaluating the predictive capacity of the model, and therefore, the risk of overfitting is higher. Statistics recommend out-of-sample validation to ensure the robustness of empirical models and avoid spurious results (Maronna et al., 2006).

5.6. Summary

The methodological analysis highlights three problems in the development of the

strategy based on volatility: the use of contemporary information in the construction of the investment signal, the choice of parameters through an intensive study of the data, and the high sensitivity to model specification. The results show the fragility of this model and indicate that results should be interpreted with caution. The analysis also underscores the importance of methodological robustness and contributes to the literature on the dangers of overfitting and data mining in finance research.

6. Discussion

6.1. Interpretation of Empirical Findings

The empirical results indicate that the volatility-based allocation strategy yields better performance compared to the Buy-and-Hold approach. Nevertheless, they demonstrate high sensitivity to parameters and model specifications. Volatility thresholds associated with superior performance are index- and time period-dependent. This implies limited parameter stability. Therefore, it cannot be assumed that observed outperformance represents an enduring structural link between past volatility and subsequent returns; rather, observed differences are likely due to sample characteristics.

6.2. Relationship to Existing Literature and Economic Interpretation

These results are partially consistent with prior studies concerning volatility timing, i.e., the notion that dynamically adjusting risk exposure can improve portfolio performance under certain conditions (Moreira & Muir, 2017). Selective elimination of specific volatility regimes appears to correspond to periods characterized by improved portfolio performance and reduced market exposure. At the same time, the findings are also consistent with prior studies on data mining and backtest overfitting in empirical finance (Bailey et al., 2015). Thus, both the high dependence of the strategy on threshold selection and the variability in results across subperiods support concerns regarding the fragility of investment strategies developed through extensive historical data exploration.

From an economic perspective, these results suggest that the relationship between volatility and expected returns is complex and unstable. Furthermore, volatility appears to correspond primarily to changing levels of uncertainty and prevailing market conditions rather than systematic variation in expected returns. Hence, relatively simple volatility-based allocation rules likely possess limited predictive power in financial markets.

6.3. Limitations and Implications for Empirical Finance

This study has two main limitations. First, the empirical analysis is confined to three equity indices without including additional asset classes or macroeconomic variables that could impact return dynamics. Second, the study does not incorporate transaction costs or other market frictions in the evaluation of strategy performance.

Overall, the results underscore the importance of methodological rigor in empirical finance research. As demonstrated in this study, quantitative investment strategies can appear highly profitable when model selection and parameter fitting rely heavily on historical data. The findings therefore reinforce the importance of robust validation procedures, out-of-sample testing, and cautious interpretation of historical backtests.

7. Conclusion

This paper examines the effectiveness of a volatility-based portfolio allocation framework for the S&P 500, DAX, and Nikkei 225 indices using data from January 1st, 1996 through December 31st, 2025. It was found that the framework produced greater cumulative and risk-adjusted returns when compared to a passive Buy-and-Hold strategy while simultaneously producing reduced market exposure during substantial portions of the sample period. Strong parameter dependence (model specification and sample period) and significant variability in the level of historical volatility associated with superior performance across indices and subperiods were additionally documented.

Further examination identified several methodological limitations that affected the interpretation of the results. Specifically, the inclusion of contemporaneous information in the $t-0$ specification of the model introduced look-ahead bias into the strategy. In addition, extensive in-sample parameter optimization together with the absence of out-of-sample validation created potentially significant risks of model overfitting. The results also indicated that the regression analysis provided limited evidence that historical volatility consistently predicts future returns. In summary, it appears that much of the apparent success of the strategy can be attributed to sample-specific characteristics and modeling decisions rather than to a stable relationship between historical volatility and future returns.

A key contribution of this study is the comprehensive review and critical evaluation of the methodology used to develop and assess a quantitative trading strategy under realistic methodological constraints. The findings further emphasize the importance of robust validation procedures and cautious interpretation of historical backtesting results in empirical finance research. More broadly, the results suggest that historical volatility may be more useful as a measure of market uncertainty and prevailing market conditions than as a reliable predictor of future returns. Future research could extend the analysis through out-of-sample testing, alternative volatility estimation methods (e.g., GARCH and EGARCH), additional asset classes, and the incorporation of transaction costs and other market frictions.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Andersen, T. (2001). The Distribution of Realized Stock Return Volatility. *Journal of Fi-*

- nancial Economics*, 61, 43-76. [https://doi.org/10.1016/s0304-405x\(01\)00055-1](https://doi.org/10.1016/s0304-405x(01)00055-1)
- Bailey, D. H., & Lopez de Prado, M. (2014). The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting and Non-Normality. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2460551>
- Bailey, D. H., Borwein, J. M., de Prado, M. L., & Zhu, Q. J. (2015). The Probability of Backtest Overfitting. *The Journal of Computational Finance*, 20, 39-69. <https://doi.org/10.21314/jcf.2016.322>
- Baillie, R. T., & DeGennaro, R. P. (1990). Stock Returns and Volatility. *The Journal of Financial and Quantitative Analysis*, 25, 203-214. <https://doi.org/10.2307/2330824>
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307-327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1998). *The Econometrics of Financial Markets*. Princeton University Press.
- Cont, R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance*, 1, 223-236. <https://doi.org/10.1080/713665670>
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1008. <https://doi.org/10.2307/1912773>
- Fleming, J., Kirby, C., & Ostdiek, B. (2003). The Economic Value of Volatility Timing Using “Realized” Volatility. *Journal of Financial Economics*, 67, 473-509. [https://doi.org/10.1016/s0304-405x\(02\)00259-3](https://doi.org/10.1016/s0304-405x(02)00259-3)
- French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected Stock Returns and Volatility. *Journal of Financial Economics*, 19, 3-29. [https://doi.org/10.1016/0304-405x\(87\)90026-2](https://doi.org/10.1016/0304-405x(87)90026-2)
- Garman, M. B., & Klass, M. J. (1980). On the Estimation of Security Price Volatilities from Historical Data. *The Journal of Business*, 53, 67-78. <https://doi.org/10.1086/296072>
- Longerstaey, J., & Spencer, M. (1996). *RiskMetricsTM—Technical Document* (4th ed.). <https://www.msci.com/documents/10199/5915b101-4206-4ba0-ae2-3449d5c7e95a>
- Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley. <https://doi.org/10.1002/0470010940>
- Moreira, A., & Muir, T. (2017). Volatility-Managed Portfolios. *The Journal of Finance*, 72, 1611-1644. <https://doi.org/10.1111/jofi.12513>
- Moreira, A., & Muir, T. (2019). Should Long-Term Investors Time Volatility? *Journal of Financial Economics*, 131, 507-527. <https://doi.org/10.1016/j.jfineco.2018.09.011>
- Parkinson, M. (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. *The Journal of Business*, 53, 61-65. <https://doi.org/10.1086/296071>

Appendix

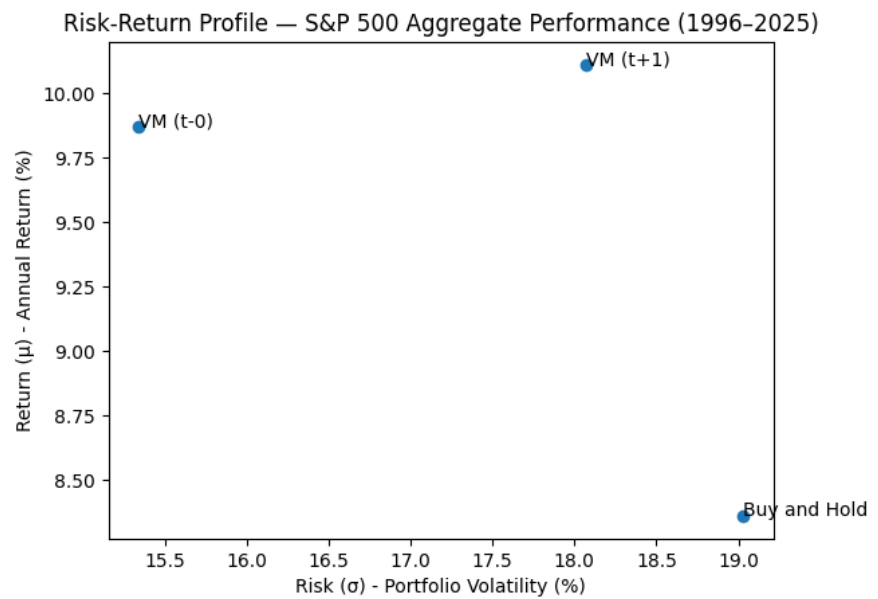


Figure A1. Risk-return tradeoff, S&P 500 (1996-2025).

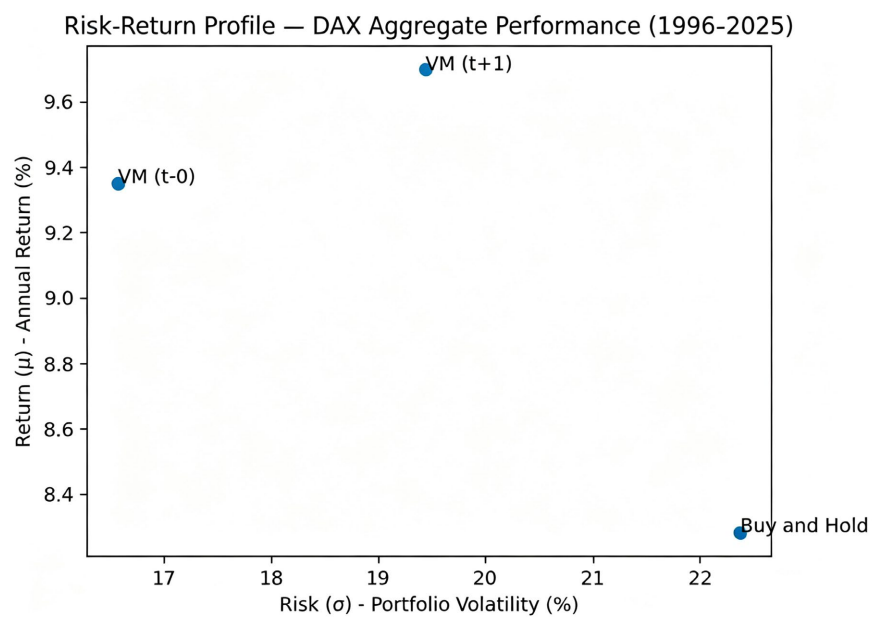


Figure A2. Risk-return tradeoff, DAX (1996-2025).

Risk-Return Profile — Nikkei 225 Aggregate Performance (1996-2025)

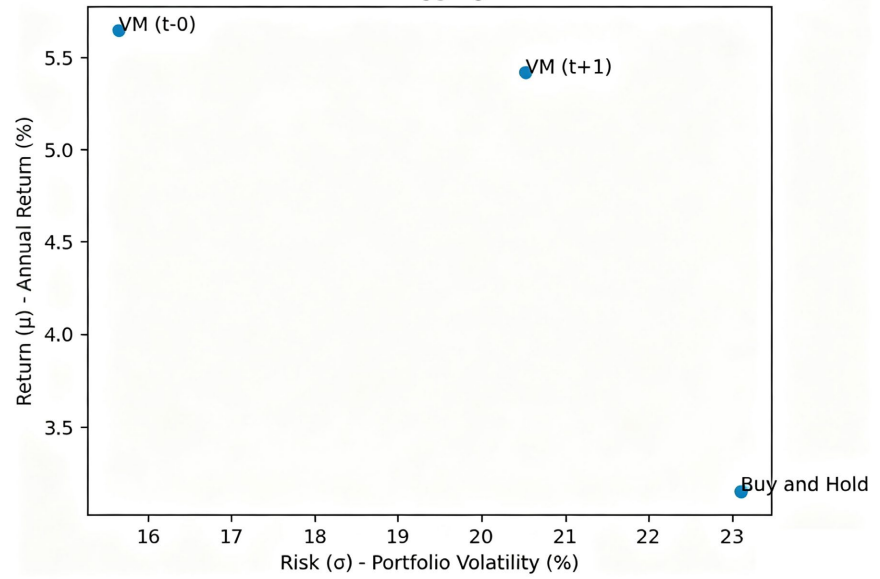


Figure A3. Risk-return tradeoff, Nikkei 225 (1996-2025), S&P 500—Subperiod Performance.

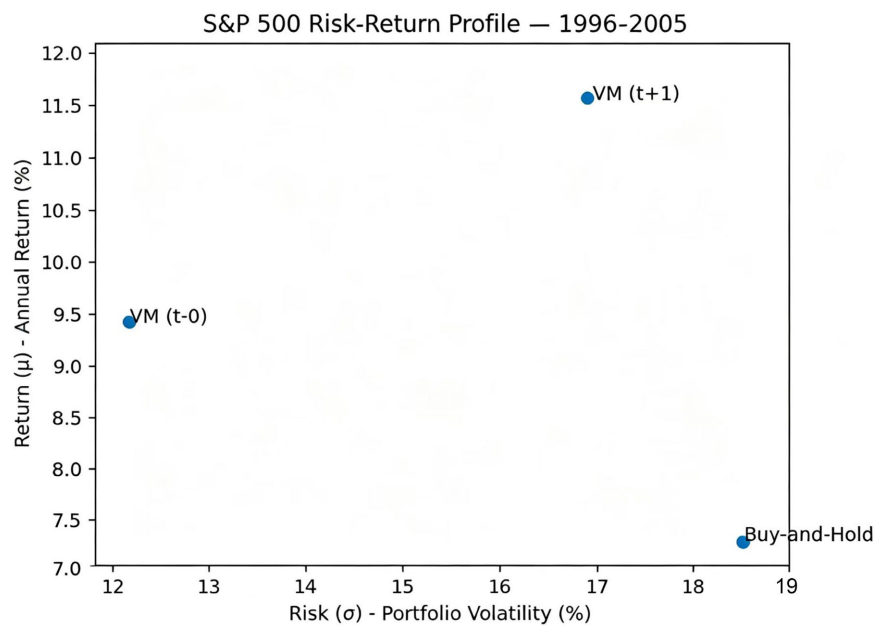


Figure A4. Risk-return tradeoff, S&P 500 (1996-2005).

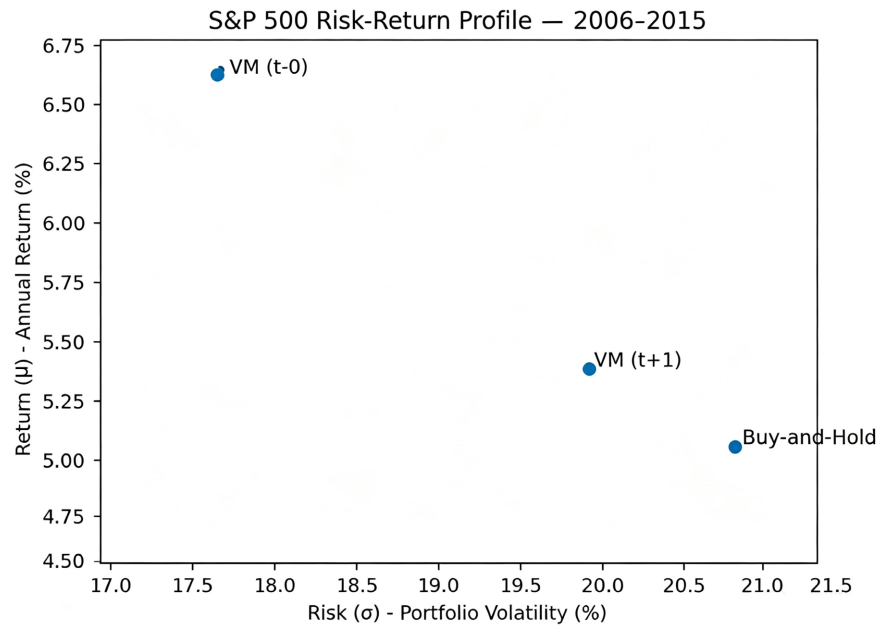


Figure A5. Risk-return tradeoff, S&P 500 (2006-2015).

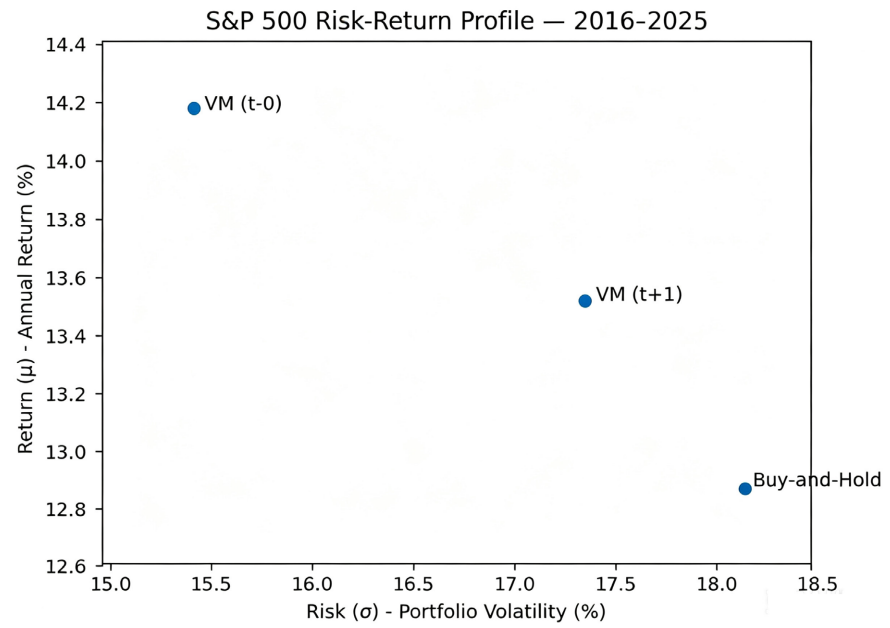


Figure A6. Risk-return tradeoff, S&P 500 (2016-2025). DAX—Subperiod Performance.

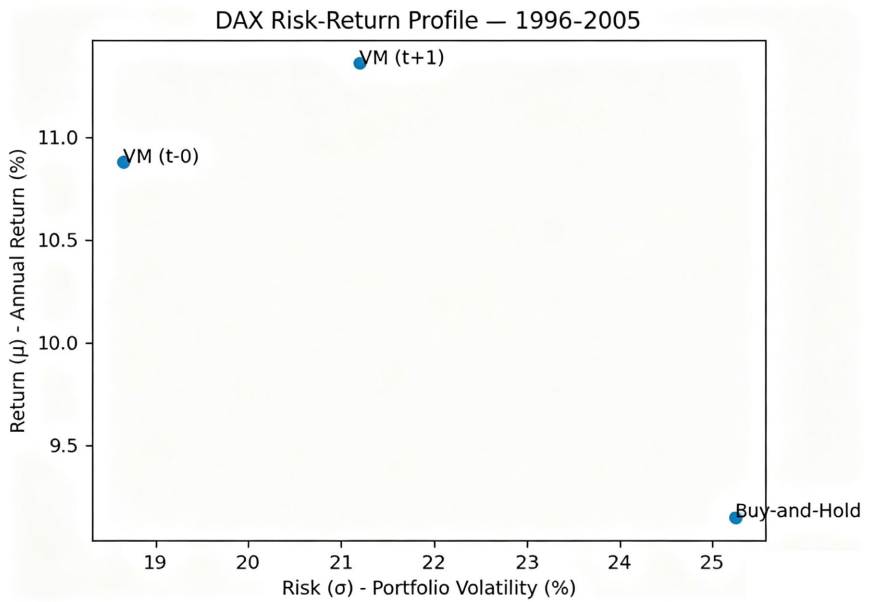


Figure A7. Risk-return tradeoff, DAX (1996-2005).

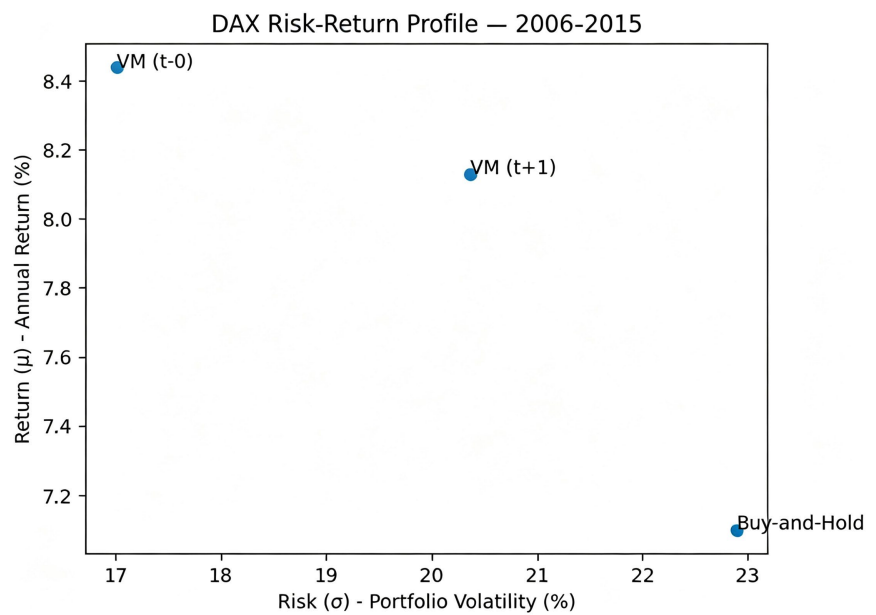


Figure A8. Risk-return tradeoff, DAX (2006-2015).

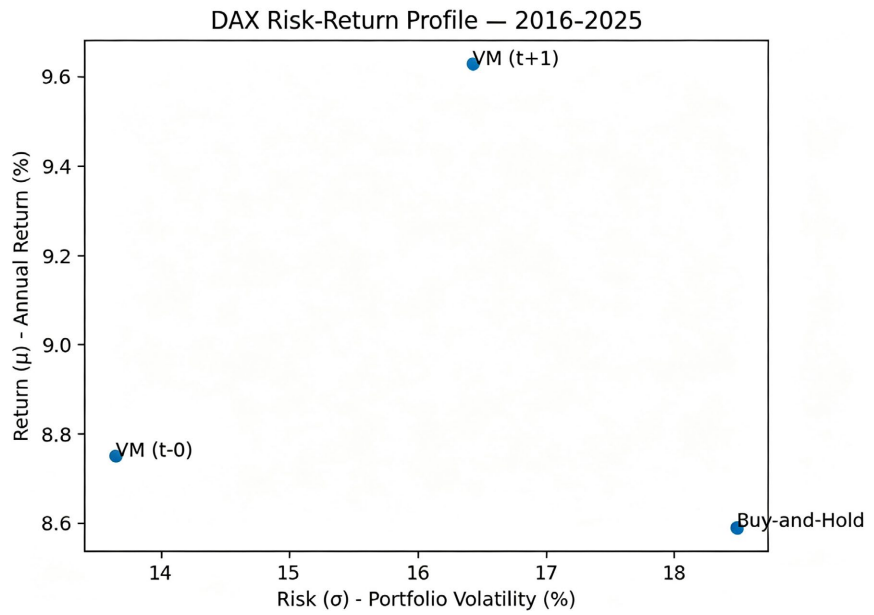


Figure A9. Risk-Return Tradeoff, DAX (2016-2025). Nikkei 225 - Subperiod Performance.

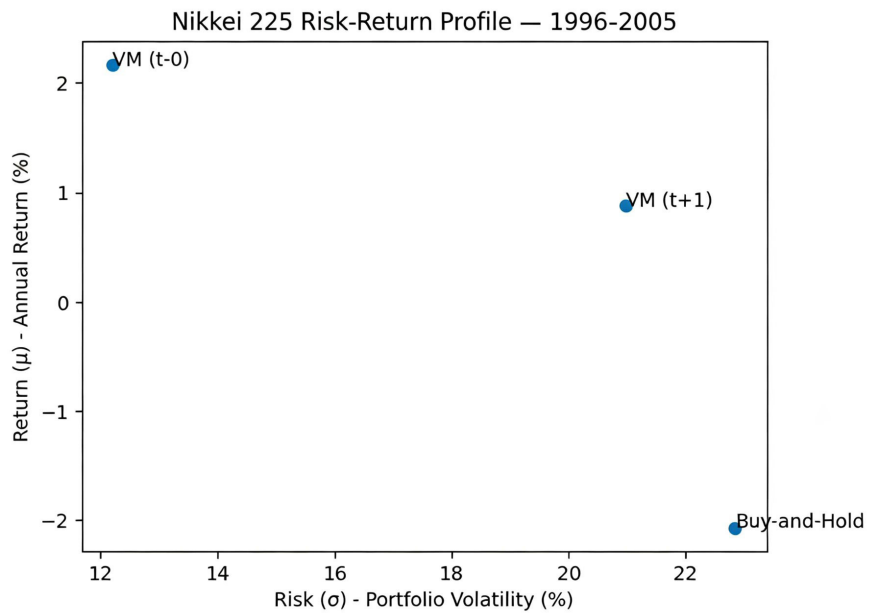


Figure A10. Risk-Return Tradeoff, Nikkei 225 (1996-2005).

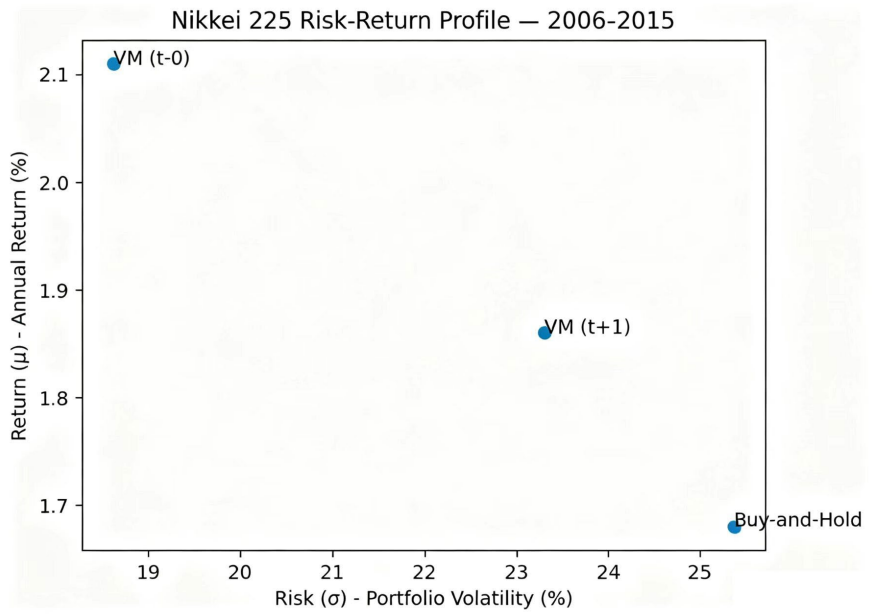


Figure A11. Risk-Return Tradeoff, Nikkei 225 (2006-2015).

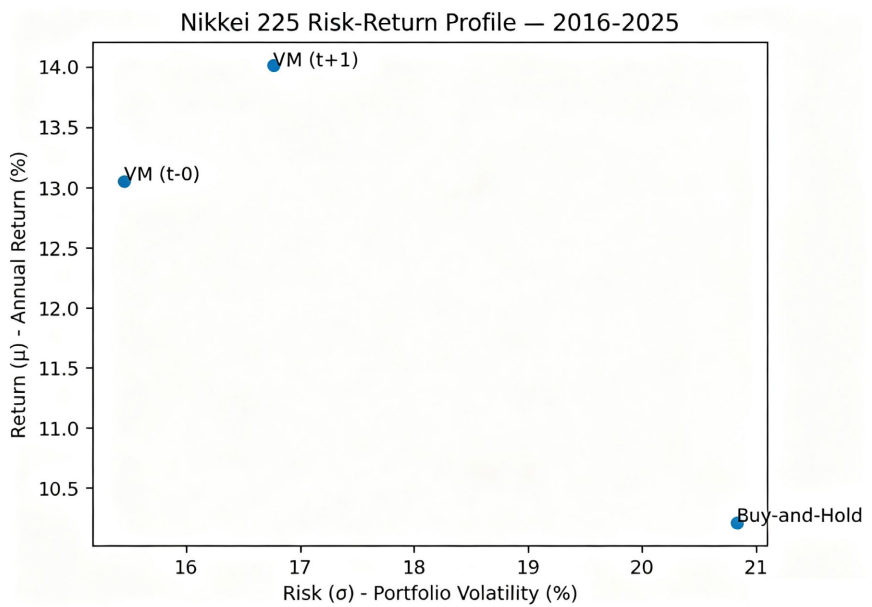


Figure A12. Risk-Return Tradeoff, Nikkei 225 (2016-2025).