國立臺灣大學管理學院財務金融研究所

碩士論文

Department of Finance
College of Management
National Taiwan University

Master’s Thesis

機器學習在波動風險溢酬策略中的應用探討

A Machine Learning Approach to Volatility Risk Premium Strategies

熊士詔
Hubert Hsiung

指導教授：蕭湛東 博士

Advisor: Lawrence Hsiao, Ph.D.

中華民國 114 年 6 月

June, 2025

doi:10.6342/NTU202503618

國立臺灣大學碩士學位論文

口試委員會審定書

機器學習在波動風險溢酬策略中的應用探討

A Machine Learning Approach to Volatility Risk Premium Strategies

本論文係熊士詔君（R12723065）在國立臺灣大學財務金融學系、所完成之碩士學位論文，於民國114年6月25日承下列考試委員審查通過及口試及格，特此證明

口試委員：

李湛中（簽名）

（指導教授）

林嘉薇

朱玉芳

系主任、所長 廖咸興（簽名）

doi:10.6342/NTU202503618

Acknowledgement

I would like to express my deepest gratitude to everyone who supported me throughout my academic journey.

To my family and friends, thank you for your unwavering encouragement, patience, and belief in me. Your support has been my foundation during both challenging and rewarding moments.

To my classmates and professors, thank you for the stimulating discussions, collaborative spirit, and academic guidance that have enriched my learning experience.

Most importantly, I would like to sincerely thank my advisor, Professor Lawrence Hsiao, for his mentorship, insightful feedback, and continuous support.

doi:10.6342/NTU202503618

摘要

本研究旨在探討機器學習技術於波動風險溢酬（Volatility Risk Premium, VRP）交易策略中的應用，透過結合行為金融與總體經濟變數，以提升策略獲利表現與穩定性。本研究應用了兩種選擇權策略：Delta-hedged short straddle 與 Delta-hedged short put 策略，並分別以「Raw VRP」與「Excess VRP」(扣除標普 500 報酬後之 VRP）為預測目標，評估多種機器學習模型之預測效能與交易表現，包括 Linear Regression、Random Forest 與 Gradient Boost 模型。本研究納入市場情緒、意見分歧、動能因子、總體經濟指標與波動率等多元變數，並透過傳統 Feature Selection 方法及 SHAP（SHapley Additive exPlanations）解釋性方法，分析模型在不同維度下之穩健性與敏感度。實證結果顯示，於兩類策略中，Delta-hedged short put 策略相對具有較高之報酬潛力與風險調整後績效；而以 Excess VRP 為目標變數之模型，整體而言具備更高之穩定性與預測能力。研究結果顯示，行為與總體經濟變數可有效強化 VRP 策略之預測性，並提供實務上進行動態部位調整之依據；同時，本研究亦揭示在市場環境變遷下維持模型穩健性所面臨之挑戰。綜上所述，本論文提出一套結合可解釋性機器學習方法與 Delta-hedged 選擇權策略之整合性架構，對於風險溢酬策略之研究與實務應用具一定貢獻。

關鍵字：波動風險溢酬、選擇權策略、機器學習、行為金融、總體經濟變數

doi:10.6342/NTU202503618

Abstract

This thesis examines the application of machine learning models to predict and optimize the profitability of volatility risk premium (VRP) strategies through adaptive delta-hedging techniques informed by behavioral and macroeconomic indicators. Specifically, we construct and evaluate delta-hedged short straddle and put strategies on the S&P 500 index. By comparing two target formulations — raw VRP and excess VRP over the underlying index — we assess the predictive performance and trading efficacy of linear regression, random forest, and gradient boosting models. Our methodology encompasses comprehensive feature engineering, incorporating sentiment, disagreement, momentum, macroeconomic surprises, and volatility structure signals. We implement both traditional and SHAP-based feature selection techniques to evaluate the sensitivity of model performance to input dimensionality. Among the two strategy types, the put-only strategy displays higher return potential and stronger Sharpe ratios. At the same time, excess VRP emerges as a more stable and predictive target variable than raw VRP. The findings demonstrate the value of incorporating behavioral finance and macroeconomic insights into quantitative models, highlighting the practical challenges of maintaining model robustness in shifting environments. This study contributes to the growing literature on machine learning applications in asset pricing by proposing a framework that combines delta-hedged options trading with interpretable predictive modeling.

Keywords: Volatility Risk Premium, Option Strategies, Machine Learning, Behavioral Finance, Macroeconomic Variables

doi:10.6342/NTU202503618

目次

doi:10.6342/NTU202503618

doi:10.6342/NTU202503618

doi:10.6342/NTU202503618

圖次

doi:10.6342/NTU202503618

doi:10.6342/NTU202503618

表次

doi:10.6342/NTU202503618

1. Introduction

1.1 Motivations

Equity investors traditionally relied on both the outperformance of individual institutions and a strong market to achieve strong returns. However, this also exposes investors to economic shocks such as the Dot-com Bubble, the Great Financial Crisis, the Euro Area Crisis, and the COVID-19 Pandemic. Due to these potential market uncertainties, investors have always sought ways to manage their market exposures, leading to the development of risk-hedging instruments such as options and swaps.

With implied volatility being a core component of financial derivatives as a measure of future risk, a growing number of researchers have directed their studies to the relationship between implied volatility and realized volatility. Jackwerth and Rubinstein (1996) found that the implied volatility for a significant drop in the S&P 500 is considerably higher than the recorded realized volatility, indicating that investors often overpay for downside protection. Empirical results suggest that buyers of hedging instruments often pay a premium to offset the risk of realized volatility exceeding implied volatility. This premium is commonly referred to as the Volatility Risk Premium (VRP), which represents the gap between implied volatility and realized volatility. Guo and Loeper (2020) further demonstrated that VRP returns are positive in the long term and can be consistently harvested. This provides market participants with a reliable method of deviating from market-associated returns in alpha (the excess return of the market) or beta (the risk relative to the market).

However, realized volatility can still exceed implied volatility at times, especially during market events that result in price jumps or sudden spikes in volatility.

doi:10.6342/NTU202503618

These scenarios often lead to significant short-term losses for VRP strategies. To address these potential shortfalls, my research aims to establish a supervised learning framework that dynamically sizes the VRP strategy positions based on behavioral and macroeconomic signals, while also identifying and evaluating key signals that contribute to the success of VRP strategies.

1.2 Contributions

This study introduces a supervised machine learning framework for dynamically allocating the position size of VRP harvesting strategies based on both behavioral and macroeconomic signals. To achieve this, it introduces and selects useful training features that improve the performance of the VRP harvesting strategy, then uses the returns of delta-hedged short put and straddle strategies as the dependent variable in my model. This research primarily focuses on Random Forest, Gradient Boosting, and Ridge Regression as models. Performance is evaluated through statistical metrics and market benchmarks. Statistical metrics include R^2 and RMSE. Market benchmarks include the Sharpe ratio, drawdown, and correlation with the S&P 500.

1.3 Paper Outline

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature on VRP and machine learning techniques with behavioral and macroeconomic signals. Section 3 provides the data, methodology, feature design, model selection, and implementation of the VRP strategy for this study. Section 4 presents the empirical results of VRP strategies within a machine learning framework, comparing them to standard VRP strategies and the S&P 500 Index. Section 5 discusses

doi:10.6342/NTU202503618

the findings and limitations of this study. Finally, section 6 concludes with suggestions for future research.

doi:10.6342/NTU202503618

2. Literature Review

2.1 Volatility Risk Premium

The volatility risk premium is commonly described as the gap between implied and realized volatility. Early research on this topic primarily focused on the inclusion of volatility in establishing option pricing models, laying the groundwork for identifying the volatility risk premium as a component of option pricing. The Black-Scholes model paved the way for future work by introducing a systematic method for option pricing (Black & Scholes, 1973). It incorporated implied volatility to represent the market's expectation of future volatility, but assumes volatility to be constant. Merton (1975) later focused on stochastic volatility and price jumps, extending the Black-Scholes model to include price movement as a component of options pricing.

Later studies focused on the inconsistency between implied and realized volatility, later identified as the volatility risk premium. French et al. (1987) investigated the relationship between the expected market risk premium and the volatility of stock returns and found the link to be positive, suggesting a larger risk premium for higher volatility. Jackwerth and Rubinstein (1996) demonstrated that the implied volatility calculated using the Black-Scholes model is often overestimated compared to the realized volatility on the downside. Christensen and Prabhala (1998) demonstrated a tendency for implied volatility to overestimate realized volatility. From this, we can reasonably conclude that hedging, especially to the downside, generally requires a premium for sellers to assume the risk of increases in realized volatility.

With empirical backing, many academics also explored the profitability of harvesting the VRP. Carr and Wu (2009) used synthetic variance swaps to trade current

doi:10.6342/NTU202503618

implied variance against future variance. Tang (2023) sold at-the-money straddles, puts, and calls on the S&P 500 index; all obtained positive average returns. However, using VRP harvesting strategies also came with significant downside risk. This risk becomes more pronounced during periods of unexpected increases in volatility, such as the Global Financial Crisis and the COVID-19 pandemic. To address this, recent research has started to look towards the predictability of VRP. Bollerslev et al. (2011) found VRP itself changes over time, suggesting it is influenced by behavioral and macroeconomic factors. My research looks at these factors and uses them as signals to improve the returns of VRP harvesting techniques.

2.2 Behavioral Effects on VRP

Since VRP changes over time, further studies have examined the factors influencing these changes. This section focuses on the potential behavioral influences on VRP predictability. Behavioral finance suggests that behavioral factors, such as investor sentiment and fear, can lead to persistent mispricing in asset markets. In the context of options, studies of behavioral effects on VRP should present evidence on the relationship between behavioral factors and implied volatility.

Smales (2014) identified a negative correlation between news sentiment and VIX, a volatility index derived from the implied volatility of SPX index options. The previously mentioned work on synthetic variance swaps by Carr and Wu (2009) attributed higher implied volatility to tail risks such as unexpected market moves. This supports the idea that investors tend to overweigh rare events due to loss aversion. Bollen and Whaley (2004) found that imbalances in order flow can influence implied volatility due to supply-demand dynamics and market frictions, resulting in a widening VRP. It is worth investigating whether the causes of the order flow imbalances are

doi:10.6342/NTU202503618

behavioral in nature, as Bollen and Whaley suggest the imbalances could stem from overreaction and demand for protection, implying VRP could be a result of emotion-driven demand. Sentiment indicators could also be helpful contributors in predicting VRP. Empirical evidence suggests that investors often overpay for downside protection during low sentiment periods, resulting in higher implied volatility and a higher VRP (Smales, 2014; Bollerslev et al., 2009). This suggests that sentiment-capturing indicators, such as volatility skew and bull-bear spreads, could potentially contribute to VRP predictability.

2.3 Macroeconomic Effects on VRP

With global economic and geopolitical risks rising in the last decade, economic uncertainties have come into focus recently. Londono et al. (2025) examined the implications of these uncertainties and found that investors demand higher risk premiums for bearing heightened risk. Empirical evidence has also shown that implied volatility increases relative to realized volatility as demand for downside protection increases. Bollerslev et al. (2009) established a connection between VRP and investor risk aversion and found correlations between risk aversion and macroeconomic variables, suggesting that macroeconomic factors may also influence VRP. Corradi et al. (2012) found VRP to be strongly countercyclical to the economy, further showing increased profitability during a pressured economy. Drechsler and Yaron (2011) focused on consumption growth and macroeconomic uncertainty to explain the time variation in VRP. These results suggest that macroeconomic-level indicators, such as GDP growth, unemployment, and inflation rates, could be incorporated into this study as predictors of VRP profitability.

doi:10.6342/NTU202503618

In addition to macroeconomic level effects, macroeconomic forecast dispersion could offer context on economic uncertainties and market disagreements. An abundance of

research already exists for the predictability of equity returns using economic forecast dispersion. For example, Bali et al. (2016) used economic dispersion as a component of uncertainty and found that equity with higher sensitivity to economic uncertainty experienced higher returns. However, similar research is lacking regarding macroeconomic forecast indicators on VRP predictability, so my study will also attempt to incorporate GDP forecasts, including forecast dispersions and surprises, into my predictive framework.

2.4 Machine Learning in Finance

The popularity of machine learning has been on the rise in finance, as machine learning models can capture complex, nonlinear relationships. This makes machine learning suitable for a wide variety of noisy financial data. Applications in finance have mostly focused on forecasting equity returns. For example, Gu et al. (2020) used Random Forest and neural network models to extract feature signals predicting stock returns, demonstrating that machine learning models can better capture nonlinearities across features compared to traditional regression techniques. Studies have also shown that machine learning better forecasts volatility than linear techniques. Christensen et al. (2022) demonstrated that machine learning models outperform linear models in predicting realized variance by comparing tree-based algorithms and neural networks with the Heterogeneous AutoRegressive (HAR) model. Bali et al. (2021) also found that non-linear machine learning models, including tree-based algorithms and neural networks, achieved higher R^2 compared to linear models when predicting option returns.

doi:10.6342/NTU202503618

However, compared to traditional econometric linear regressions, the "black box" nature of machine learning models makes feature interpretation harder to infer. As a result, feature selection and model interpretability become important areas of study, especially in highly regulated fields such as finance. Rudin (2018) argues that machine learning models should generally prioritize interpretability in high-stakes domains such as finance, since erroneous interpretations could result in detrimental outcomes, and a lack of interpretation could also become an obstacle to accountability. Guyon and Elisseeff (2003) provided an overview of feature selection methods. This paper follows their guidance and uses feature importance and permutation importance as feature selection methods. The study also employs SHAP (SHapley Additive exPlanations) values, as SHAP has the ability to provide both local and global explanations across complex non-linear models (Lundberg & Lee, 2017).

doi:10.6342/NTU202503618

3. Data & Methodologies

3.1 Behavioral & Economic Features

3.1.1 VIX

The VIX index represents the 30-day implied volatility of the S&P 500. As a forward-looking indicator derived from SPX option prices, VIX captures investor expectations of future volatility and is widely used as a proxy for market uncertainty. Bollerslev et al. (2011) show that the VIX can predict the VRP since it represents the implied volatility used in option-based variance swaps. Smales (2014) also notes that the VIX serves as a fear gauge, linking it to behavioral factors that influence risk premia. Due to its theoretical and empirical significance, the VIX is regarded as a key variable in predicting VRP.

3.1.2 Trading Volume

Trading volume serves as an important indicator of market activity. To pinpoint periods of unusually high or low investor participation, we use both a rolling z-score of volume, which standardizes the raw numbers against their average and standard deviation, and the standard raw trading volume. Typically, periods of high trading volume coincide with increased uncertainty or the arrival of new information, which influences option pricing and implied volatility (Chordia et al., 2000). Research by Bali et al. (2021) suggests that abnormal trading volume, particularly when combined with sentiment signals, enhances our ability to predict option returns. Furthermore, volume often correlates with limits-to-arbitrage and imbalances in order flow, which can shift VRP due to changes in option demand (Bollen & Whaley, 2004).

doi:10.6342/NTU202503618

3.1.3 Relative Strength Index (RSI)

RSI is a tool that helps traders gauge the momentum of price movements. It identifies whether the market is overbought or oversold, providing insights into investor sentiment or potential market fatigue. Research by Go et al. (2020) shows that RSI is commonly integrated into predictive equity models. Its effectiveness increases when combined with nonlinear features in machine learning, enabling the model to identify reversal and continuation trends, which aids in option pricing and understanding implied volatility.

3.1.4 Put-Call Skew

Put-call skew, which highlights the difference in implied volatility between out-of-the-money puts and calls, indicates the asymmetries in option demand. It shows how much investors lean towards seeking downside protection and can serve as a gauge for crash risk or the likelihood of rare adverse events. Bates (2000) views volatility skew as a measure of perceived left-tail risk tied to investor behavior, while Bollen and Whaley (2004) reveal that shifts in skew are influenced by demand imbalances. This can distort option prices and increase VRP. Thus, skew functions as both a sentiment indicator and a reflection of risk aversion among investors.

3.1.5 Momentum (1M, 3M, 6M, 12M)

We look at cumulative returns for 1, 3, 6, and 12-month periods to understand how price momentum works. Research has shown that previous returns can help predict future returns because investors often underreact or take time to adjust prices (Jegadeesh & Titman, 1993). In the realm of options, Bali et al. (2021) demonstrate that return momentum can indicate option returns and errors in volatility pricing. By examining

doi:10.6342/NTU202503618

various time horizons, we can better identify short-term and long-term behavioral trends related to VRP. To include trading intensity in momentum measures, we also calculated volume-weighted momentum using rolling averages of price multiplied by volume. This method reflects the strength behind price changes. Bali et al. (2016) suggest that momentum signals adjusted for volume provide deeper insights into investor sentiment and differing opinions, which can help spot times when the VRP is mispriced.

3.1.6 Price Gap Ratio

The price gap ratio measures the difference between the current day's opening price and the previous day's closing price, adjusted for the prior closing price. This metric serves as an indicator of overnight changes in investor sentiment. Research conducted by Lou et al. (2019) shows that these overnight returns are influenced by how investors react to news released after market hours, often leading to behavioral biases such as overreaction. By integrating this metric, the model considers external information sources outside of regular trading hours that could affect adjustments in implied volatility.

3.1.7 Intraday Range

Intraday range is commonly indicated by the difference between the highest and lowest prices during the day, measured against the closing price. Garman and Klass (1980) demonstrated that these high-low ranges effectively estimate daily volatility. This helps us better understand daily implied volatility levels and the volatility risk premium.

doi:10.6342/NTU202503618

3.1.8 Unemployment Rate

The unemployment rate is an important indicator of the overall health of the labor market. In the context of risk premia, studies by Drechsler and Yaron (2011) and Corradi et al. (2012) suggest that implied volatility tends to rise during periods of economic distress, as indicated by increased unemployment rates. As a result, the unemployment rate reflects broader economic conditions and how investors perceive the risk associated with them.

3.1.9 CPI MoM

Monthly changes in the Consumer Price Index (CPI) are a key indicator of inflation. Unexpected shifts in inflation and overall uncertainty can influence monetary policy expectations, which in turn affect equity market volatility and VRP. Bollerslev et al. (2009) demonstrated that both macroeconomic uncertainty and risk aversion play a role in the changing nature of the VRP, reinforcing the importance of including CPI as a predictive macroeconomic variable.

3.1.10 GDP Forecast/Surprise/Dispersion

GDP forecasts share insights into expected economic growth, typically sourced from consensus surveys. GDP surprise is calculated by taking the difference between actual GDP growth and what was predicted by analysts. When unexpected changes in economic growth occur, they can lead to increased volatility, prompting investors to adjust their expectations for future market fluctuations. Research by Andersen et al. (2003) shows that macroeconomic surprises often correlate with sudden changes in asset prices and volatility, highlighting the importance of incorporating these surprises into models that focus on volatility. Economic uncertainty and disagreement among

doi:10.6342/NTU202503618

economists can be understood through forecast dispersion, which is measured by the standard deviation of GDP forecasts. Research conducted by Bloom (2009) and Bali et al. (2016) indicates that these dispersion indicators reflect informational frictions and perceived risks. These factors are crucial for grasping volatility pricing.

3.1.11 Bull-Bear Spread

The AAII Bull-Bear Spread reflects the difference between bullish and bearish sentiment among investors. According to Baker and Wurgler (2007), this spread is linked to extreme market sentiments, which can lead to mispricing. Smales (2014) further demonstrates that indicators driven by sentiment often correlate with the VIX and implied volatility. For this reason, the Bull-Bear Spread is included to highlight the behavioral aspects of investor fear and exuberance, which can affect the dynamics of the VRP.

3.2 Strategy Construction

This study employs strategies to harvest VRP using delta-hedged short options positions on the S&P 500 index. It focuses on two main strategies: the delta-hedged short straddle and the delta-hedged short put. Both strategies aim to exploit the difference between implied and realized volatility while reducing directional market exposure through daily delta-hedging. This approach is consistent with the methodology of Tang (2023), who demonstrated that delta-hedged short straddles and puts on the S&P 500 tend to yield excess returns over time due to the ongoing pricing discrepancy between expected and realized volatility.

There are three essential steps to implementing these strategies: (1) initiating the strategy at the start of each month, (2) performing daily mark-to-market valuations and

doi:10.6342/NTU202503618

delta hedge adjustments, and (3) rolling to new positions at monthly expiration. Each strategy is based on a 30-calendar-day European-style option on the S&P 500 index. The actual price movements are derived from historical daily levels of the S&P 500 index. Option prices are determined using the Black-Scholes model, with implied volatility derived from the VIX, while the risk-free rate is set at the 10-year U.S. Treasury yield. At the start of every month, the strategy involves selling options in amounts relative to the current NAV. To balance out directional exposure, the portfolio is delta-hedged daily. This hedge is achieved by synthetically simulating a position in the SPX futures that offsets the net delta of the sold options.

The daily profit and loss of the strategy includes three elements: (1) the changes in option value (option P&L), (2) the profit or loss from using the underlying index for hedging (hedging P&L), and (3) the interest accrued on any unused cash (cash interest P&L). The strategy position is adjusted daily and rolled at the end of the month, when a new position is established.

3.2.1 Delta-Hedged Short Straddle

The short straddle strategy entails selling an at-the-money call and put option at the same strike price and expiration date. This approach seeks to isolate the return generated from volatility mispricing by eliminating market exposure.

3.2.2 Delta-Hedged Short Put

In the second strategy, we focus on selling only at-the-money put options. This approach still retains a directional element, as investors primarily buy puts for downside protection. Similar to straddles, the pricing and hedging of short puts involve daily delta hedging to minimize directional risk. This strategy highlights a different aspect of the

doi:10.6342/NTU202503618

VRP as the demand for puts often indicates investor anxiety and the need for protection against extreme market movements. Compared to the straddle, the short put strategy generally shows a stronger correlation with equity market returns and can provide more appealing risk-adjusted returns during stable market conditions.

3.2.3 Dynamic Position Sizing

This study presents a dynamic sizing framework based on supervised learning to improve risk management and potentially enhance performance. Each month, we estimate the expected return of the VRP strategy. This predicted value is then converted into a position weight, adjusting the size of the short options position accordingly. The transformation function is designed to increase position size when anticipated returns are high and to decrease size, or even significantly reduce it, when expected returns are low or negative. The definition of this weighting function is as follows:

$$w_t = \begin{cases} \min(1 + 100 * \hat{y}_t, maximum\_weight), & \hat{y}_t \geq 0 \\ \max\left(\frac{1}{-100 * \hat{y}_t + 1}, minimum\_weight\right), & \hat{y}_t < 0 \end{cases}$$

$\hat{y}_t$ represents the expected return, with the minimum and maximum weights set at 0.2 and 5, respectively, indicating that the weight limits are set at a factor of 5. This scaling ensures symmetry between positive and negative predictions while avoiding excessive leverage.

3.2.4 Time Range

The training period spans from January 1990 to December 2014, enabling models to learn from over 20 years of macroeconomic changes, behavioral signals, and market volatility cycles. This extensive historical dataset helps the model to understand different market situations, including recessions, expansions, and crises.

doi:10.6342/NTU202503618

Following the training set, a validation period from January 2015 to December 2017 is used to assess the model's generalizability and to tune the hyperparameters.

Finally, a test period from January 2018 to June 2023 is used to evaluate the performance of the models. This timeframe covers significant events such as the start of deglobalization, the COVID-19 market shock, rising inflation concerns, and shifts in monetary policy, making it a valuable time to assess the model performance.

3.2.5 NAV and Performance Tracking

The strategy is assessed by examining the cumulative NAV, which begins at $ 100 million USD and fluctuates with daily profits and losses. To evaluate the effectiveness of the machine learning approach, model-driven VRP strategies with dynamic scaling are compared to standard static VRP strategies and the S&P 500 index. Performance is measured both in statistical terms, such as RMSE and R^2, and through risk-adjusted metrics, including the Sharpe ratio and maximum drawdown. These comparisons reveal whether predictive signals enhance the risk-return profile of the strategy compared to a basic volatility carry trade.

3.3 Target Variable: VRP Profitability

A central component of this study is creating a target variable to train supervised learning models that predict the profitability of VRP harvesting strategies. The VRP is typically defined as the gap between implied volatility in option prices and the actual realized volatility of the underlying asset. However, VRP strategies aim to profit from this premium in real-life applications through structured trades. Therefore, the actual returns from these strategies provide a better representation of VRP profitability than the volatility spread itself. To implement this, we define the target variable as the monthly

doi:10.6342/NTU202503618

profitability of delta-hedged short options strategies through two evaluation methods. These evaluations are crucial for training our models, assessing their performance, and determining whether machine learning can reliably predict returns based on VRP.

3.3.1 Method 1: Absolute Strategy Returns

The first method assesses the profitability of VRP by examining the absolute returns of delta-hedged options strategies. In this approach, we compute the monthly changes in NAV for each strategy and calculate monthly returns using the following formula:

$$R_t^{strategy} = \frac{NAV_t - NAV_{t-1}}{NAV_{t-1}}$$

where NAV_t is the end-of-month NAV of the strategy, and NAV_{t-1} is the end-of-month NAV of the previous month. This approach indicates the profit or loss of the VRP strategy, making it suitable for assessing the ability of machine learning models to predict times of high performance or increased risk.

Absolute returns can provide useful information, but it is essential to consider the broader market conditions' impact. For instance, when the equity market is performing well, it can skew VRP strategies upward due to hedging profits and losses from our imperfect hedging, even if implied volatility is not significantly high. Hence, we may need a different perspective to control for beta.

3.3.2 Method 2: Excess Returns over the S&P 500

To address this concern, the second method assesses the profitability of VRP by comparing its excess return to that of the S&P 500 index. This follows the approach that Guo and Loeper (2020) and Bollerslev et al. (2011) outlined. By benchmarking the

doi:10.6342/NTU202503618

strategy’s monthly return against the S&P 500's contemporaneous return, this method effectively controls for beta and directional exposure:

$$R_t^{excess} = R_t^{strategy} - R_t^{SPX}$$

This excess return highlights the strategy’s alpha in relation to overall market performance and focuses on the risk premium gained from volatility mispricing, without relying on equity beta. This is particularly significant in the context of VRP harvesting, as strategies can achieve returns even when equity markets are flat or declining. Additionally, adhering to the widely accepted definition of alpha in asset pricing facilitates direct comparisons with traditional benchmarks and factor models.

3.4 Feature Engineering

To ensure that the predictive signals are empirically robust, all features in this study are designed to reflect the information available at the time a trade is initiated. Since the VRP harvesting strategies operate monthly and options are priced over a 30-day period, features are built using past daily observations and then resampled to a monthly frequency. This approach prevents any lookahead bias from affecting the training process. By combining behavioral and macroeconomic signals into a cohesive dataset with proper lag structure and resampling, this feature engineering process guarantees that all inputs are practical and properly timed with the VRP strategy horizon.

Many behavioral and market variables, such as momentum and sentiment indicators, are available daily. To align with the option trade execution window, which starts at the beginning of each month, these daily indicators are adjusted by one day and then resampled to reflect end-of-month values. This process ensures that the model only uses information that would have been available before initiating a new straddle.

doi:10.6342/NTU202503618

Macroeconomic indicators such as unemployment rates, GDP, and inflation are typically reported on a monthly or quarterly basis. These figures can either be joined directly on their release dates or estimated through forward-fill methods when necessary.

3.5 Model Selection

3.5.1 Ridge Regression

Ridge Regression serves as the basic linear model in this study. It is a modified version of Ordinary Least Squares (OLS) regression that applies an L2 penalty to help decrease overfitting by reducing the magnitude of the coefficients.

3.5.2 Random Forest

Random Forest is an ensemble learning method that constructs multiple decision trees using bootstrapped samples and random subsets of features. It predicts outcomes by averaging the results from the individual trees, which helps reduce variance and improve generalization. This method is especially useful for financial prediction tasks because it can: (1) capture nonlinear relationships between input variables and targets, (2) effectively deal with multicollinearity, and (3) measure feature importance through impurity reduction. In this study, Random Forest is implemented using scikit-learn's RandomForestRegressor, with key hyperparameters (number of trees, tree depth, minimum samples per leaf) tuned via grid search on the validation set.

doi:10.6342/NTU202503618

3.5.3 Gradient Boosting

Gradient Boosting is a type of tree-based ensemble method that constructs trees sequentially to correct the residuals of previous ones. This allows GBMs to achieve higher predictive accuracy, but at the cost of increased complexity and susceptibility to overfitting. This study utilizes the HistGradientBoostingRegressor from scikit-learn, which is known for its speed and its native support for missing values. Key hyperparameters, including learning rate, number of estimators, and maximum tree depth, are tuned via grid search on the validation set.

3.5.4 Rationale for Model Selection

The selection of these three models represents a balance between simplicity, interpretability, and predictive performance. Ridge Regression is straightforward and acts as a baseline linear model. Random Forest strikes a good balance, being both interpretable and powerful, allowing for the identification of important features and their interactions. Gradient Boosting, while providing top-tier accuracy, requires more intricate tuning and can be more challenging to interpret.

This study uses a combination of models to evaluate both the top-performing model in terms of predictions and the features that are consistently significant across different frameworks. The variety in modeling approaches also helps determine whether any performance enhancements are specific to one model or applicable across various types of algorithms.

doi:10.6342/NTU202503618

3.6 Feature Selection

3.6.1 Tree-Based Models: Random Forest and Gradient Boosting

For the Random Forest and Gradient Boosting models, we used three methods for feature selection: mean decrease impurity (MDI) feature importance, permutation importance, and SHAP values.

Feature importance scores from tree-based models give us an overall idea of how frequently and effectively each feature contributes to splitting decision nodes. These scores indicate the reduction in variance linked to each feature, averaged across all trees. Features that score higher in importance are seen as having a greater impact on the predictions made by the model.

However, MDI-based feature importances can be biased when features or variables are correlated or have many unique values. To address this issue, the study also uses permutation importance. This method assesses the contribution of each feature by looking at how the model's performance changes (measured using out-of-sample R^2 and RMSE) when the feature's values are randomly shuffled. Unlike MDI, permutation importance evaluates the impact of each feature relative to the overall trained model and is compatible with any model type.

Beyond feature importance metrics, SHAP values enhance interpretability. It provides both local and global insights into feature influence, assigning a value to each feature based on its contribution to predictions across various combinations. This method allows for a detailed understanding of which features consistently lead to higher or lower VRP profitability and how their impacts change over time and in different market conditions. SHAP values are particularly effective for tree-based models since the TreeSHAP algorithm ensures precise computation of feature contributions.

doi:10.6342/NTU202503618

Additionally, we created a pairwise feature correlation heatmap to identify multicollinearity and any redundant variables. Although tree-based models tend to handle multicollinearity well due to their hierarchical splitting structures, significant correlations can still hinder interpretability and inflate variable importance scores. The heatmap helped us qualitatively assess whether specific features should be excluded or combined.

3.6.2 Ridge Regression

Unlike tree-based models, Ridge Regression is a linear method that employs L2 regularization. It tends to shrink the coefficients of correlated features instead of removing them entirely. This characteristic helps reduce both overfitting and multicollinearity, making Ridge Regression a suitable choice for handling noisy financial data. Alpha values are selected using RidgeCV. However, ridge regression lacks built-in features for interpreting the importance of features beyond simply examining the size of the coefficients. As a result, this study did not utilize advanced techniques for feature selection, such as permutation importance or SHAP, in conjunction with Ridge Regression.

To address this limitation, we conducted a residual analysis to check for any systematic bias or heteroskedasticity in the errors of the linear model. This diagnostic step was important for determining whether we were overlooking relevant nonlinearities and whether we might need to consider feature transformations or interaction terms to improve the model in the future.

doi:10.6342/NTU202503618

4. Empirical Results

4.1 Overview of Experimental Setup

This study evaluates how behavioral and macroeconomic indicators can predict the profitability of VRP using a machine learning approach. We use three different models: Ridge Regression, Random Forest, and Gradient Boosting. Each model forecasts monthly VRP profitability, which is then used to dynamically size positions in delta-hedged short volatility strategies. To measure relative performance, the results are compared with the S&P 500 Index. Each model is trained to predict either: (1) Absolute VRP profitability (defined as the monthly return of the delta-hedged strategy), or (2) Excess VRP profitability (defined as the strategy return minus the S&P 500 return for the same period). The research uses data ranging from January 1990 to June 2023. The dataset is split into three separate periods: a training set from 1990 to 2014, a validation set from 2015 to 2017, and a testing set from 2018 to 2023.

The strategy involves monthly adjustments to either a delta-hedged short put or a delta-hedged short straddle position on the S&P 500 index. To estimate option prices, we utilize the Black-Scholes model, with the VIX serving as a proxy for implied volatility and a 30-day maturity period. Position sizing is performed using a dynamic scaling function based on the expected VRP profitability. The portfolio NAV is updated daily to account for option market values, delta hedging profit and loss, and interest earned on cash.

Feature selection for tree-based models is informed by feature importance scores, permutation importance, and SHAP values, with correlated variables monitored using a correlation heatmap. For Ridge Regression, residual diagnostics are employed to

doi:10.6342/NTU202503618

evaluate predictive performance, though multicollinearity is not explicitly corrected for beyond standardization. Hyperparameter tuning for the Random Forest and Gradient Boosting models is performed via grid search on the validation set. For Ridge Regression, alpha values are selected via RidgeCV.

All models are evaluated using multiple metrics, including statistical metrics such as R-squared and RMSE, as well as performance metrics like annualized returns, Sharpe ratio, and maximum drawdowns, to ensure a robust assessment across both statistical and economic dimensions.

Metric	**Baseline Straddle**	**Baseline Put**	**SPX**
Ann. Return	0.83%	3.99%	10.97%
Sharpe Ratio	-0.05	0.06	0.14
Max Drawdown	-0.19	-0.19	-0.25
Win Rate	64.06%	68.75%	67.19%
SPX Correlation	0.4	0.86	1

Table 4.1 Baseline Strategy and SPX Metrics

doi:10.6342/NTU202503618

Model	Strategy	Target Variable (VRP)	Feature Set	R²	RMSE	Ann. Return (%)	ML Sharpe	ML Max DD	ML Win Rate (%)	SPX Corr
Linear	Straddle	Raw	All	-0.2303	0.00046	-0.92	-0.07	-0.3	64.06	0.39
Linear	Straddle	Excess	All	0.2294	0.00138	1.32	-0.02	-0.23	65.62	0.1
Linear	Put	Raw	All	0.5937	0.00033	0.46	-0.02	-0.39	68.75	0.66
Linear	Put	Excess	All	-0.052	0.00088	14.98	0.29	-0.16	70.31	0.8
RF	Straddle	Raw	All	-4.5252	0.01588	0.78	-0.04	-0.28	64.06	0.4
RF	Straddle	Raw	Top 10	0.5263	0.02292	0.14	-0.05	-0.26	65.62	0.19
RF	Straddle	Raw	Top 5	0.6326	0.02308	-1.32	-0.09	-0.27	64.06	0.1
RF	Straddle	Excess	All	0.3404	0.02605	-0.02	-0.06	-0.26	65.62	0.23
RF	Straddle	Excess	Top 10	0.5169	0.02585	-2.46	-0.11	-0.3	64.06	0.17
RF	Straddle	Excess	Top 5	0.6721	0.02248	-1.32	-0.08	-0.27	64.06	0.08
RF	Put	Raw	All	-0.4486	0.0177	1	-0.02	-0.35	68.75	0.7
RF	Put	Raw	Top 10	0.3378	0.0145	0.39	-0.03	-0.36	68.75	0.69
RF	Put	Raw	Top 5	0.6607	0.01153	0.48	-0.02	-0.38	68.75	0.69
RF	Put	Excess	All	-0.1442	0.02012	7.29	0.13	-0.17	70.31	0.81
RF	Put	Excess	Top 10	0.4945	0.01754	9.55	0.15	-0.21	70.31	0.83
RF	Put	Excess	Top 5	0.691	0.01502	10.51	0.18	-0.16	70.31	0.84
GB	Straddle	Raw	All	0.4153	0.01477	0.66	-0.04	-0.3	65.62	0.39
GB	Straddle	Raw	Top 10	0.4789	0.01413	-0.16	-0.06	-0.28	65.62	0.44
GB	Straddle	Raw	Top 5	0.5174	0.0136	0.61	-0.04	-0.25	65.62	0.43
GB	Straddle	Excess	All	0.6117	0.02639	-0.4	-0.06	-0.29	65.62	0.23
GB	Straddle	Excess	Top 10	0.7868	0.02303	-1.92	-0.1	-0.29	64.06	0.18
GB	Straddle	Excess	Top 5	0.7938	0.02209	-1.76	-0.1	-0.28	64.06	0.11
GB	Put	Raw	All	0.6621	0.01649	1.06	-0.02	-0.35	68.75	0.71
GB	Put	Raw	Top 10	0.676	0.01527	-1	-0.05	-0.41	68.75	0.68
GB	Put	Raw	Top 5	0.8299	0.01113	-0.13	-0.03	-0.39	68.75	0.69
GB	Put	Excess	All	0.6046	0.01816	7.38	0.13	-0.18	70.31	0.78
GB	Put	Excess	Top 10	0.7609	0.01644	9.76	0.17	-0.18	70.31	0.84
GB	Put	Excess	Top 5	0.7469	0.01612	10.22	0.17	-0.19	70.31	0.82

Table 4.2 Model Performance Metrics Comparison

doi:10.6342/NTU202503618

4.2 Ridge Regression Results

4.2.1 Straddle Strategy

VRP as Target Variable

The Ridge Regression model trained to predict raw VRP from delta-hedged short straddles performed poorly on the test set, with an R^2 of -0.23 and an RMSE of 0.00046. As shown in Figure 4.1, the predicted returns failed to track the realized VRP profitably, resulting in ML performance worse than that of the baseline strategy and the market. The model's Sharpe ratio (-0.07) and annualized return (-0.92%) highlight its limited practical value. The residual plot in Figure 4.2 indicates that the predicted values were systematically off-target, particularly at turning points in volatility cycles.

The model's coefficients (Table 4.3) suggest that positive VRP is weakly associated with short-term momentum indicators and macroeconomic signals. Specifically, 1-month momentum, unemployment, and GDP surprise had the strongest positive weights, while VIX, RSI, and Volume_Z were negatively associated with future VRP. However, the small absolute magnitude of most coefficients and the low overall R^2 imply that the model lacked sufficient explanatory power to generalize well out of sample.

doi:10.6342/NTU202503618

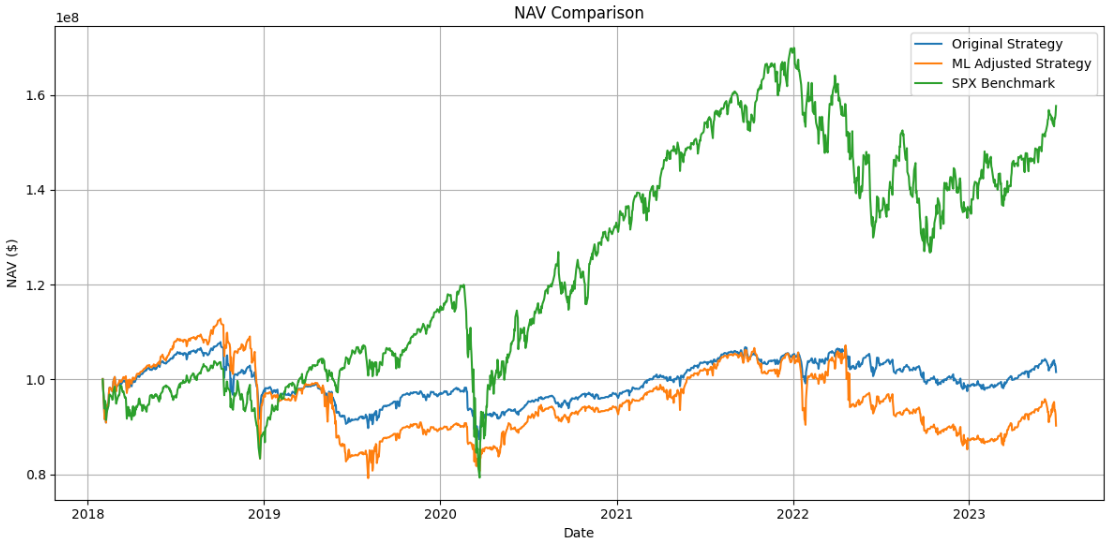

Fig. 4.1 Performance (NAV) comparison for Ridge Regression using straddle strategy and VRP as the target

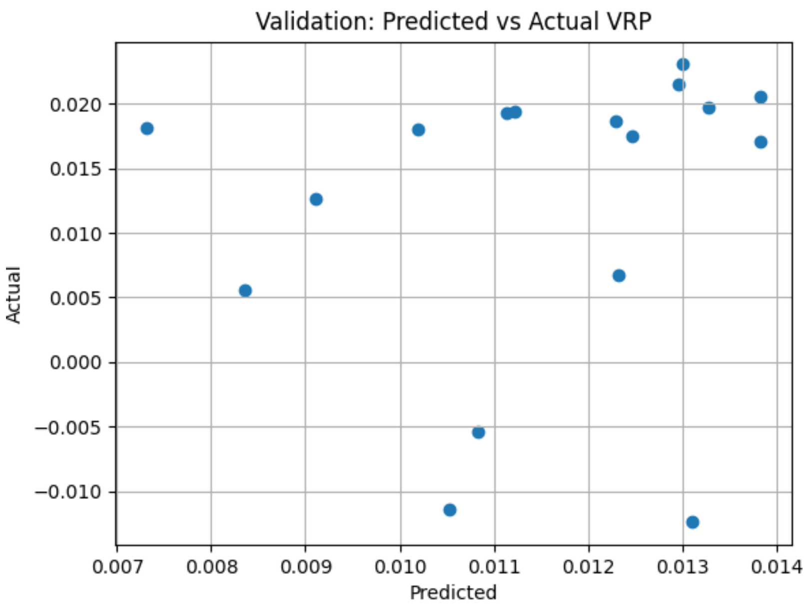

Fig. 4.2 Residuals for Ridge Regression using straddle strategy and VRP as the target

doi:10.6342/NTU202503618

Feature	Straddle (Raw VRP)	Straddle (Excess VRP)	Put (Raw VRP)	Put (Excess VRP)
Momentum_1M	0.004306	-0.03027	0.010973	-0.02342
Unemployment	0.001901	0.001962	0.000744	0.000631
PutCallSkew	0.001203	0.001482	0.000669	0.00079
GDP_Surprise	0.000621	-0.0005	0.000574	-0.00055
Momentum_12M	0.000418	-0.000085	0.00072	0.000357
GDP_Forecast	0.000039	0.000847	-0.00009	0.000742
PriceGapRatio	-0.00012	-0.00208	0.000374	-0.00157
IntradayRange	-0.00028	0.002536	-0.000003	0.002721
GDP_Dispersion	-0.00029	-0.00039	-0.00017	-0.00027
Momentum_6M	-0.00056	-0.00057	-0.00079	-0.00096
CPI_MoM	-0.00056	-0.0003	-0.00056	-0.00025
Bull_Bear_Spread	-0.00083	-0.0014	0.000232	-0.00032
Volume_Z	-0.00096	-0.00301	0.000214	-0.00179
Momentum_3M	-0.00129	-0.00278	0.000121	-0.00137
RSI	-0.00252	-0.00562	0.005188	0.002378
VIX	-0.00314	-0.00702	-0.00083	-0.00446
Momentum_1M	0.004306	-0.03027	0.010973	-0.02342
Unemployment	0.001901	0.001962	0.000744	0.000631
PutCallSkew	0.001203	0.001482	0.000669	0.00079
GDP_Surprise	0.000621	-0.0005	0.000574	-0.00055
Momentum_12M	0.000418	-0.000085	0.00072	0.000357

Table 4.3 Ridge Regression Feature Coefficients

doi:10.6342/NTU202503618

Excess VRP as Target Variable

When targeting excess VRP, model performance improved notably, with an R^2 of 0.229 and RMSE of 0.00138. This model, shown in Figure 4.3, captured periods of relative VRP outperformance more effectively, yielding a modest improvement in annualized return (1.32%) and win rate (65.6%). The residuals in Figure 4.4 exhibit a more dispersed but less biased prediction structure than the raw VRP case.

Analyzing the coefficients reveals a shift in the relevance of features. Intraday range, unemployment, and Put-Call skew emerged as top contributors to excess VRP, while momentum indicators, especially short- and mid-term momentum, carried strong negative weights. Interestingly, VIX and RSI also displayed large negative coefficients, reinforcing the notion that excess VRP tends to decline during periods of heightened volatility and overbought conditions.

This dichotomy may reflect how raw VRP responds more to market direction (momentum), whereas excess VRP better isolates the volatility premium from market beta. Behavioral features, such as Put-Call skew and Bull-Bear spread, remain relevant, indicating that investor sentiment and protection-seeking behavior can influence the pricing discrepancy between implied and realized volatility beyond simple market trends.

doi:10.6342/NTU202503618

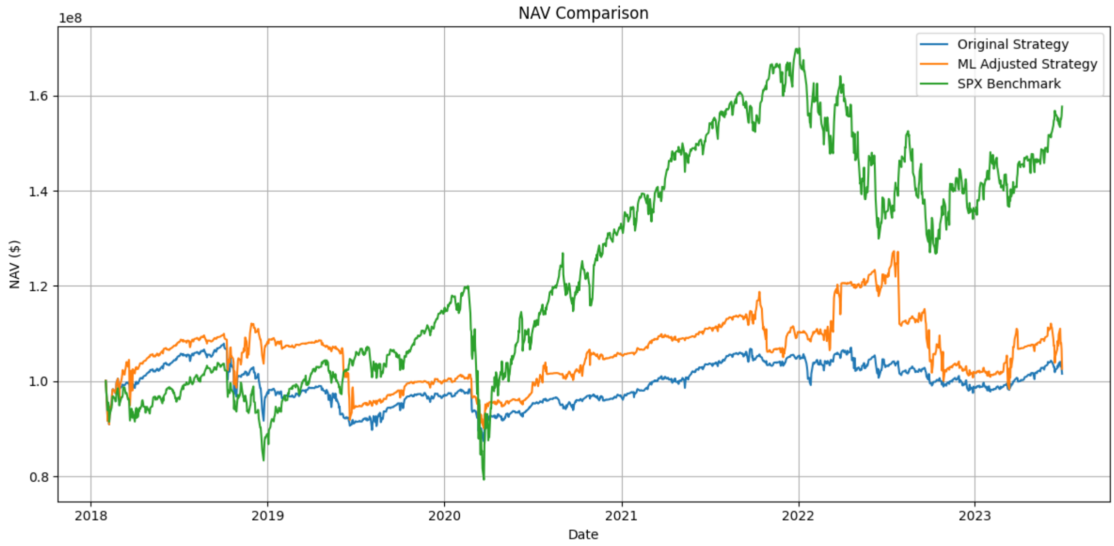

Fig. 4.3 Performance (NAV) comparison for Ridge Regression using straddle strategy and excess VRP as the target

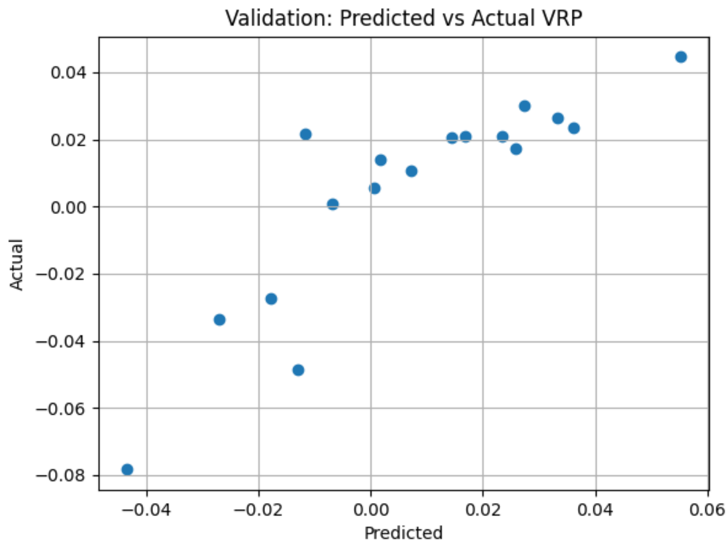

Fig. 4.4 Residuals for Ridge Regression using straddle strategy and excess VRP as the target

doi:10.6342/NTU202503618

4.2.2 Put Strategy

VRP as Target Variable

The put-only strategy with raw VRP as the target variable achieved the highest test-set R^2 (0.594) among all Ridge Regression experiments and relatively low RMSE (0.00033). However, Figure 4.5 shows the model did not translate into a strong trading performance, Figure 4.6 reveals tighter clustering around zero, indicating better prediction alignment with the observed VRP.

Table 4.3 indicates that Momentum_1M, RSI, and Unemployment are the dominant positive predictors. Several behavioral variables, including Put-Call skew, Bull-Bear spread, and Volume_Z, also had modest positive contributions. The strong weight on Momentum_1M suggests short-term price trends positively inform VRP, possibly due to autocorrelation in volatility responses to market rallies or drops. The limited contribution of macro variables, such as GDP Forecasts or CPI, suggests that the model primarily exploited behavioral or technical structure.

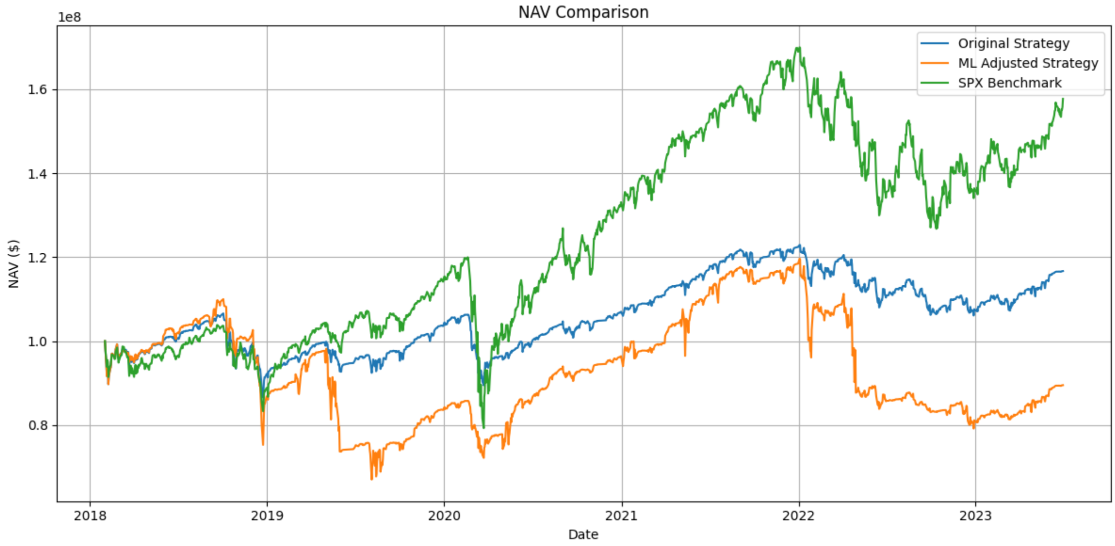

Fig. 4.5 Performance (NAV) comparison for Ridge Regression using put strategy and raw VRP as the target

doi:10.6342/NTU202503618

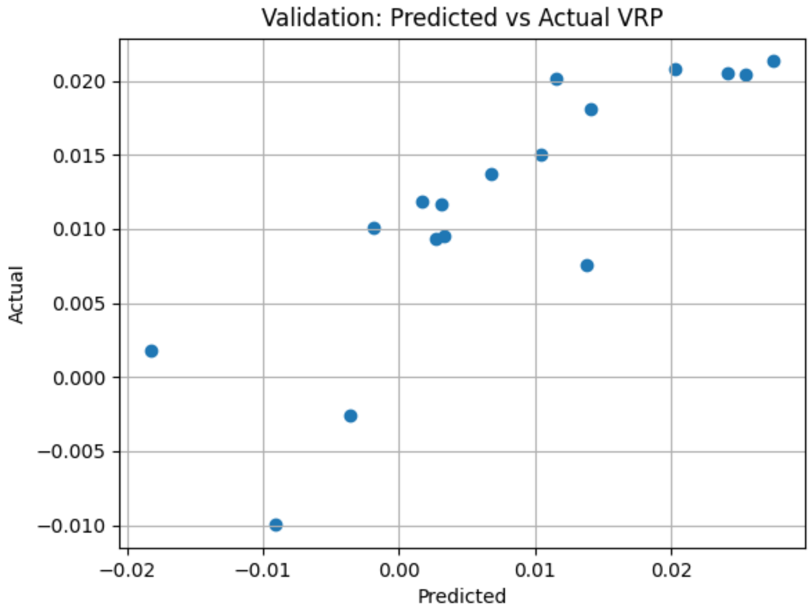

Fig. 4.6 Residuals for Ridge Regression using the put strategy and raw VRP as the target

Excess VRP as Target Variable

Switching the dependent variable to excess VRP reduced predictive performance (R^2 = -0.052), but notably improved market performance, with a Sharpe ratio of 0.29 and an annualized return of 14.98%, outperforming both the baseline (3.99%) and the SPX (10.97%). Figure 4.7 illustrates the smoother and more profitable NAV trajectory resulting from this model. Interestingly, despite the lower R^2, the strategy captures a consistent directional signal.

In this case, Intraday range, RSI, and Put-Call skew once again dominate the positive coefficient rankings, while Momentum_1M, VIX, and Volume_Z receive large negative weights. This reinforces the finding that excess VRP is more related to volatility asymmetry and behavioral positioning than to trend-following momentum. In contrast to the raw VRP target, macroeconomic indicators play a minor role, suggesting

doi:10.6342/NTU202503618

that outperformance is better explained by market microstructure effects rather than fundamental changes.

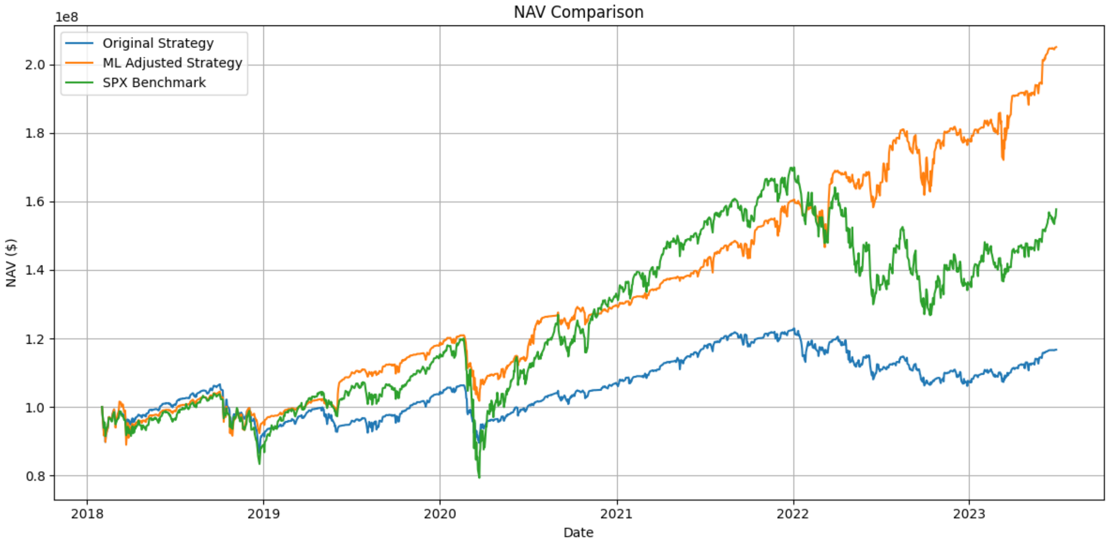

Fig. 4.7 Performance (NAV) comparison for Ridge Regression using put strategy and excess VRP as the target

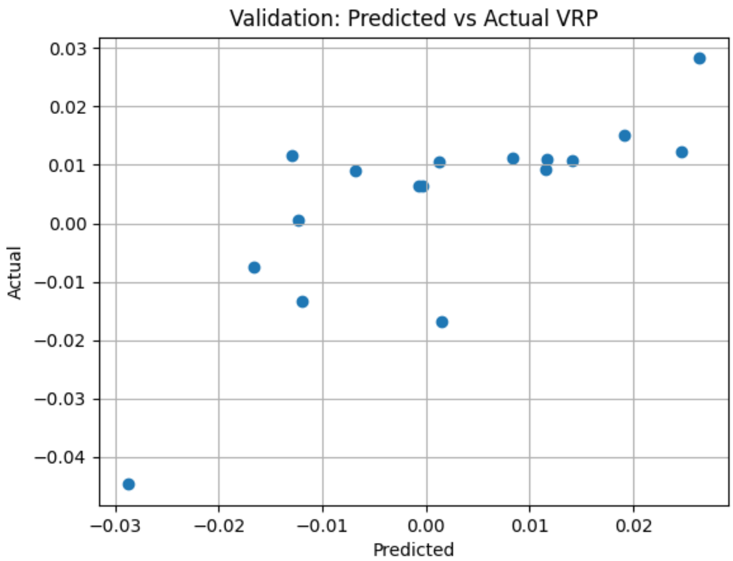

Fig. 4.8 Residuals for Ridge Regression using the put strategy and raw VRP as the target

doi:10.6342/NTU202503618

4.3 Random Forest Results

4.3.1 Straddle Strategy

VRP as Target Variable

When trained to predict raw VRP, the Random Forest model displayed significant improvements in fit and interpretability when feature dimensionality was constrained. Using all features, the model produced a poor R^2 of -4.53 and RMSE of 0.0159, indicating considerable overfitting or noise sensitivity. However, once limited to the top 10 SHAP-ranked features, the R^2 improved to 0.53 with a more reasonable RMSE of 0.0229. Using only the top 5 features yielded a slightly lower R^2 of 0.63, indicating that most of the explanatory power is concentrated in a small subset of inputs.

The model's predicted VRP-based portfolio exhibited meaningful variation. Using all features, the annualized return was 0.78%. In comparison, the top 10 feature set yielded 0.14%, and the top 5 resulted in a negative return of -1.32%. Sharpe ratios similarly declined from -0.04 to -0.09 as dimensionality decreased. Figure 4.9 illustrates the NAV trajectories of these models, while Figure 4.10 provides comparative SHAP value plots.

doi:10.6342/NTU202503618

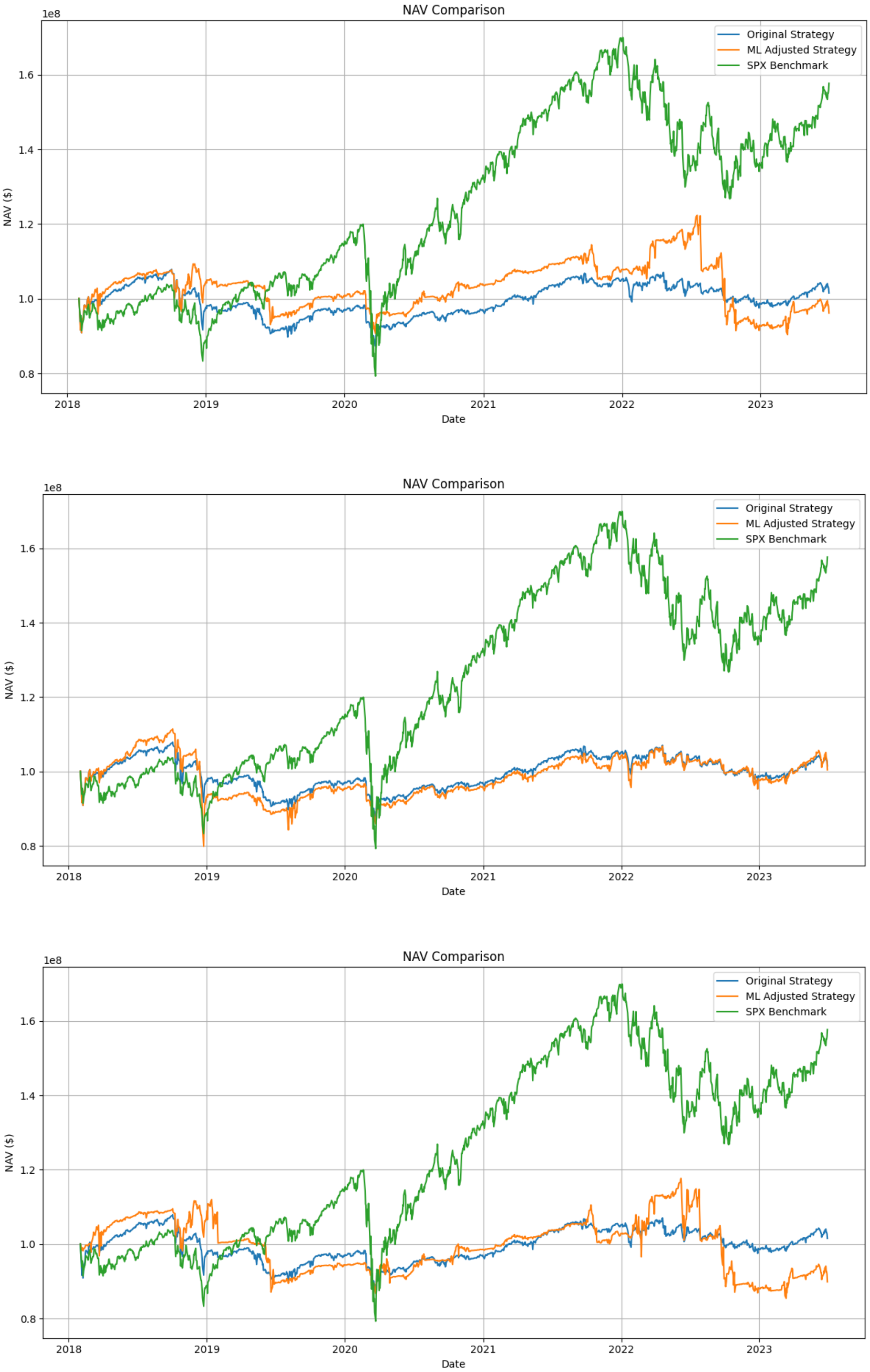

Fig. 4.9 Performance (NAV) comparison for Random Forest using straddle strategy and raw VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

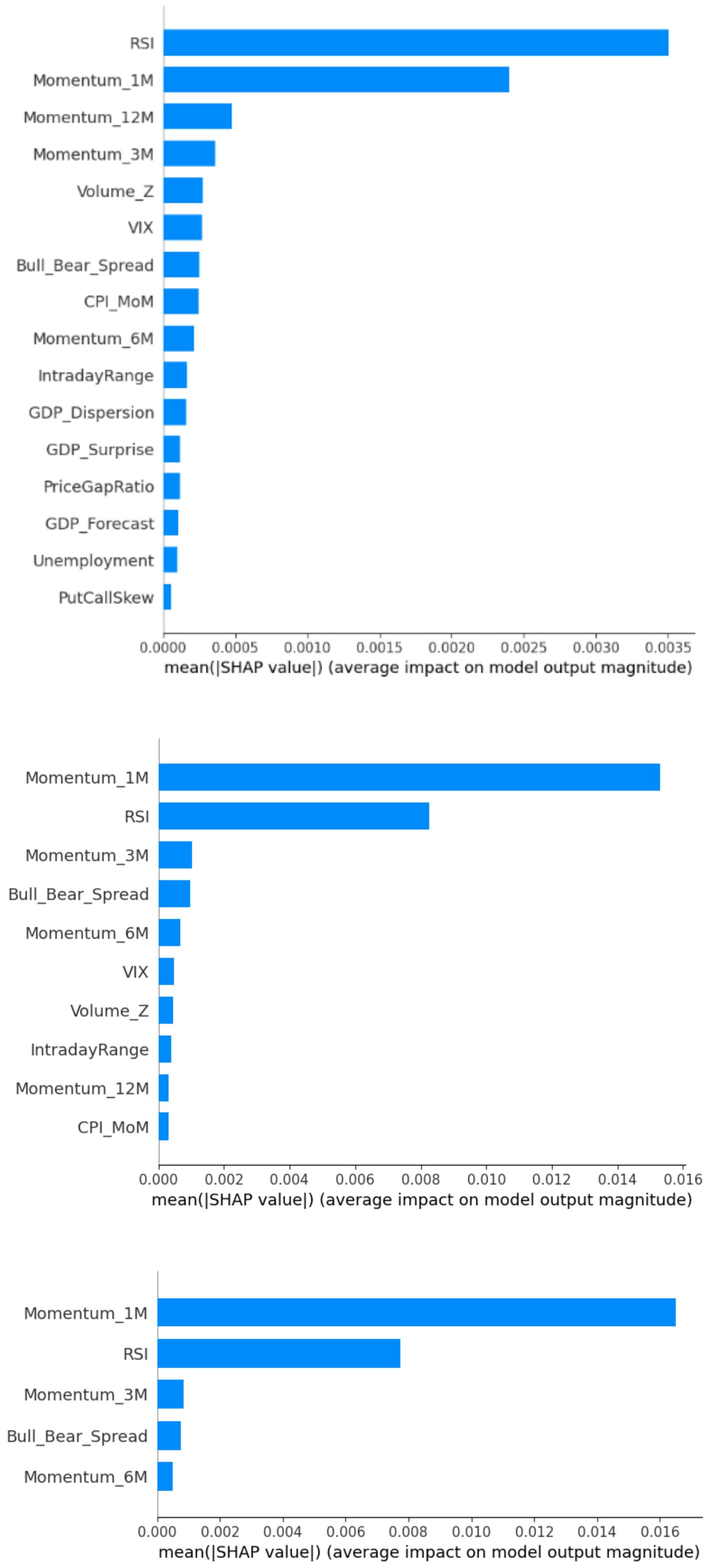

Fig. 4.10 SHAP value comparison for Random Forest using straddle strategy and raw VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

Excess VRP as Target Variable

When targeting excess VRP, the Random Forest model demonstrated notably better statistical and trading performance. Using all features, the model achieved an R^2 of 0.34 and an RMSE of 0.0260. Limiting the input to the top 10 features improved the R^2 to 0.52, while a top 5 subset achieved the highest explanatory power with an R^2 of 0.67 and a reduced RMSE of 0.0225.

Despite improvements in fit, the model-driven dynamic scaling strategy exhibited underperformance: the ML-based portfolio delivered returns of -0.02%, -2.46%, and -1.32%, respectively, for all, top 10, and top 5 feature sets. These negative returns suggest that, while the model was statistically better at predicting excess VRP, this did not translate into superior trading outcomes.

Figures 4.11 and 4.12 present the NAV and SHAP comparisons, respectively, revealing consistent dominance of behavioral and sentiment-driven indicators such as RSI, Momentum_1M, and Bull-Bear Spread. These features appear recurrently across both target specifications, suggesting their importance for predicting volatility risk premiums in straddle strategies.

doi:10.6342/NTU202503618

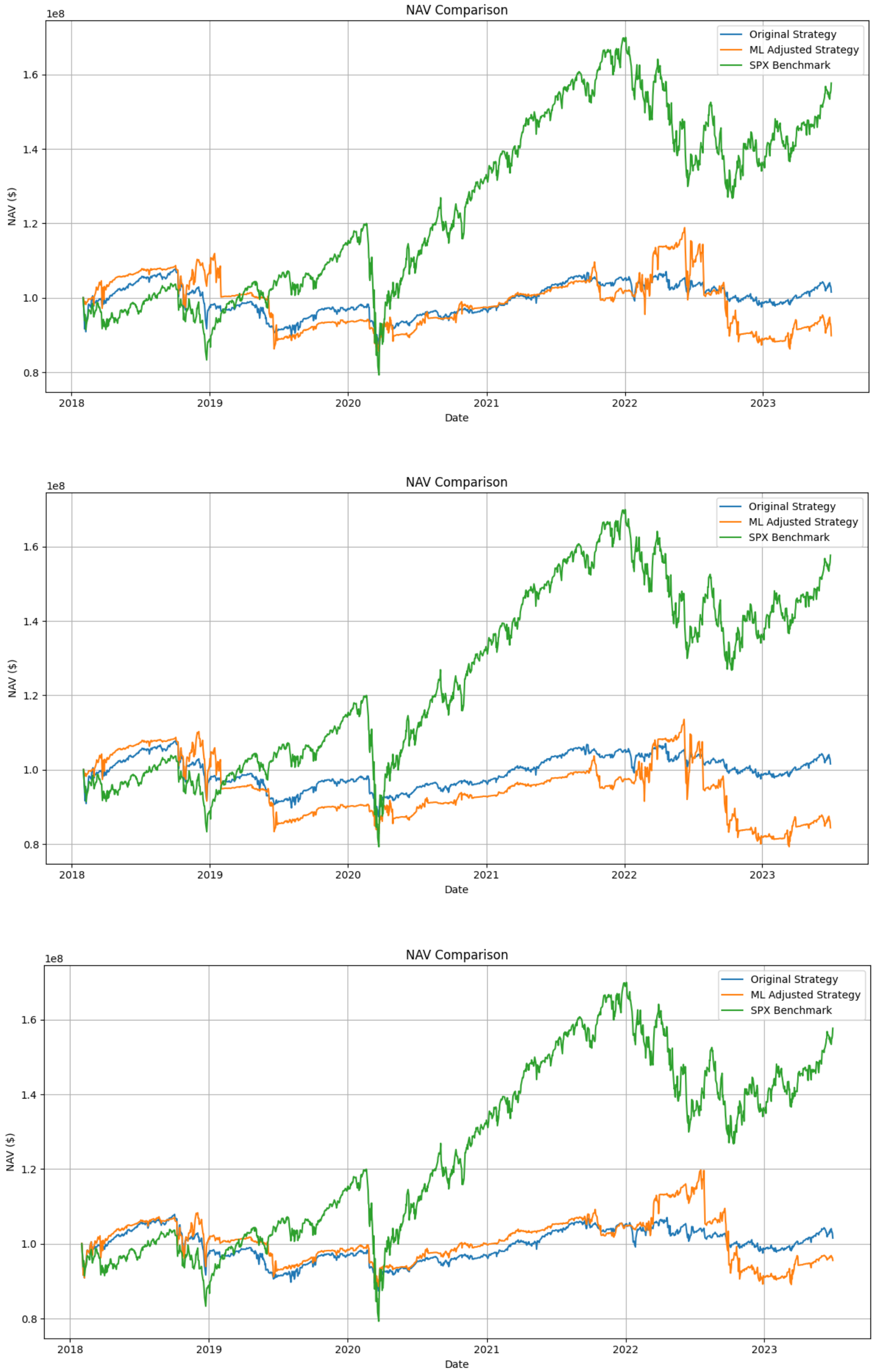

Fig. 4.11 Performance (NAV) comparison for Random Forest using straddle strategy and excess VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

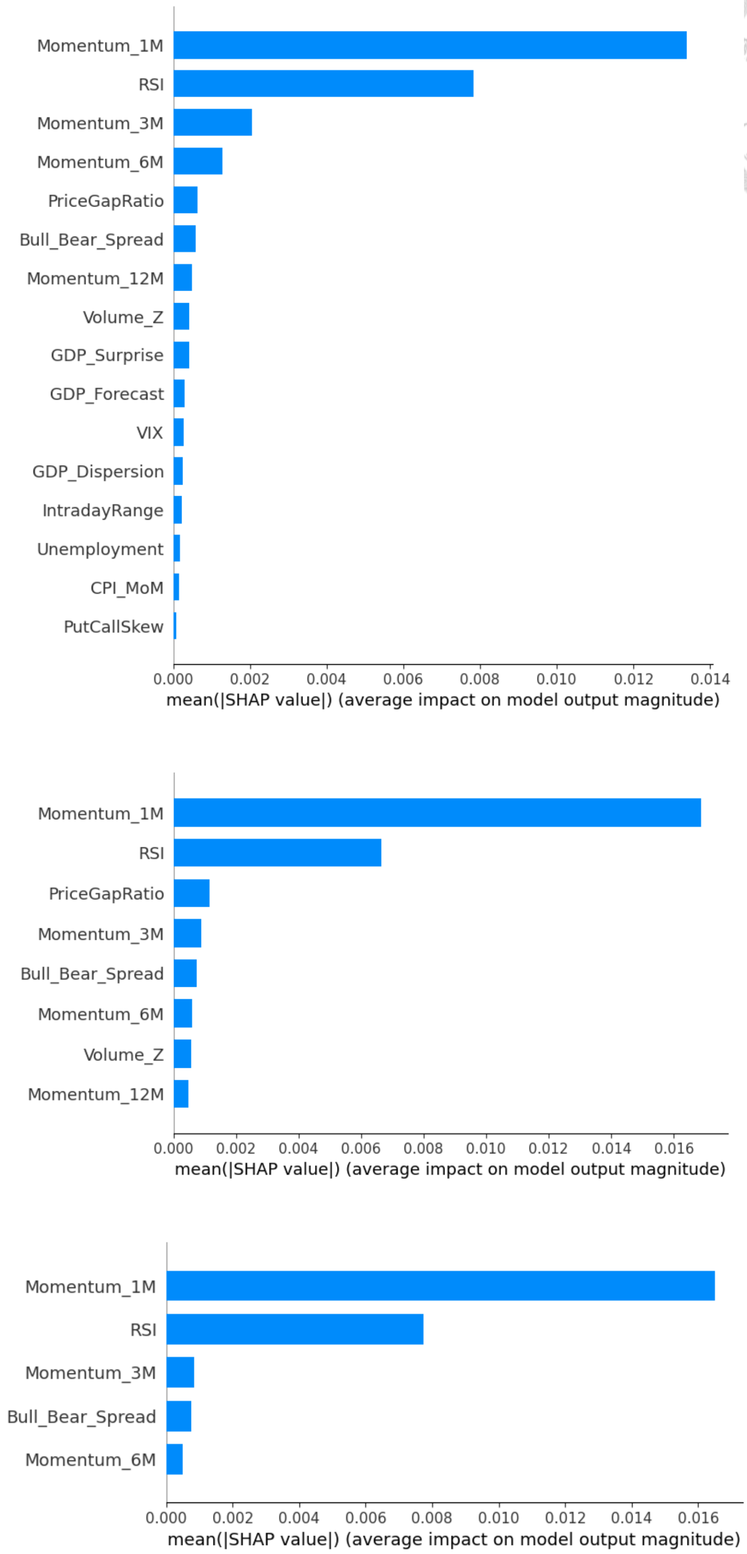

Fig. 4.12 SHAP value comparison for Random Forest using straddle strategy and excess VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

4.3.2 Put Strategy

VRP as Target Variable

The Random Forest model performed modestly when applied to the put strategy with raw VRP as the target. With all features included, the R^2 was negative (-0.45), and RMSE was 0.0177, indicating overfitting and poor generalization. However, model performance improved considerably as the feature set was pruned. The top 10 features yielded an R^2 of 0.34, while the top 5 features increased the R^2 to 0.66 and reduced RMSE to 0.0115.

Market-level annualized returns remained relatively stable at 1% for the all-feature model, while the top 10 and top 5 sets yielded returns of 0.39% and 0.48%, respectively. The win rate was 68.75% across the board. Sharpe ratios remained negative throughout, likely reflecting high variance and suboptimal timing in harvesting the volatility premium.

SHAP values visualized in Figure 4.14 demonstrate that macroeconomic and volatility-related indicators such as CPI_MoM, GDP Forecasts, and VIX were particularly influential in raw VRP prediction. Performance NAV comparisons in Figure 4.13 also reflect the model's inability to significantly enhance base strategy performance, despite achieving a higher statistical fit in the reduced feature sets.

doi:10.6342/NTU202503618

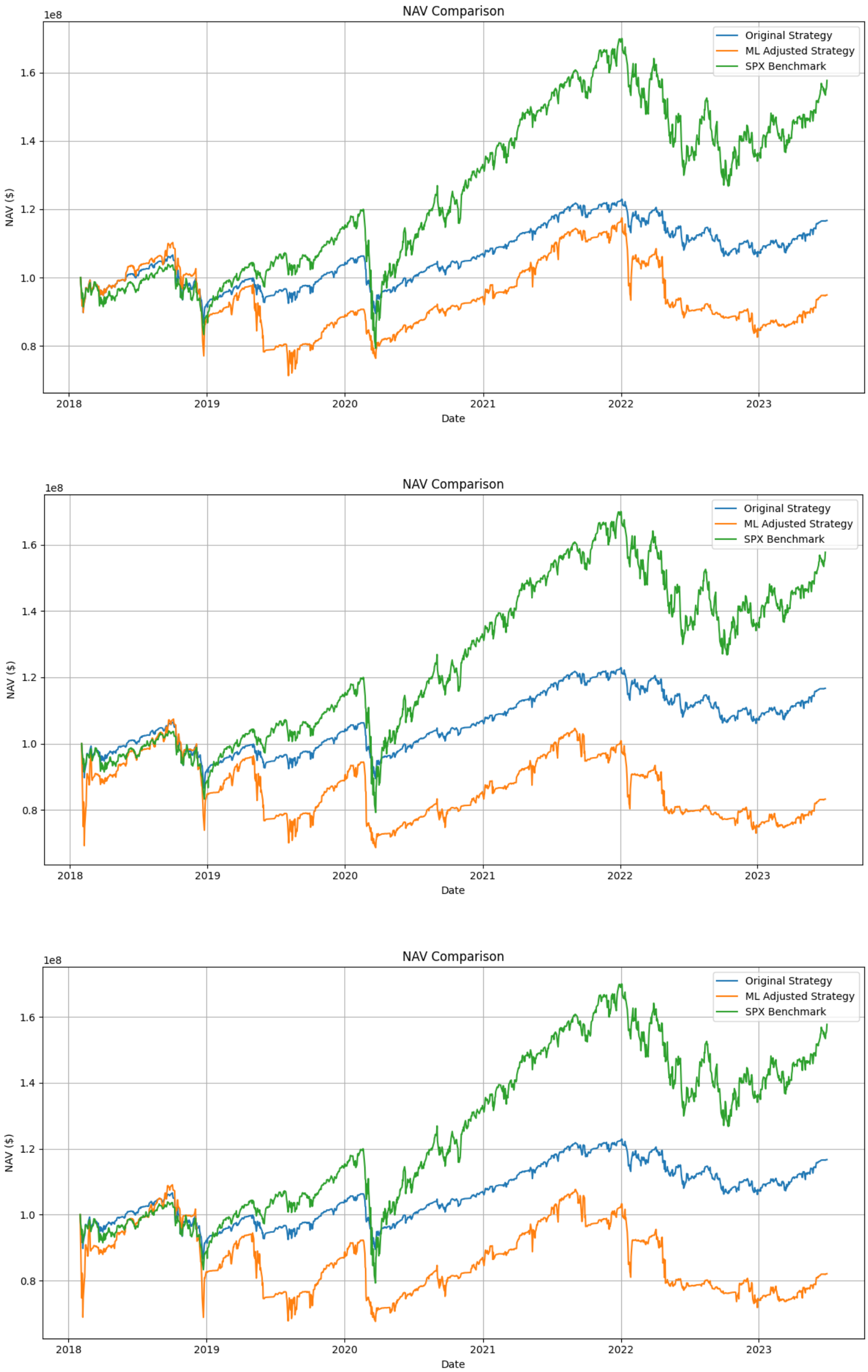

Fig. 4.13 Performance (NAV) comparison for Random Forest using put strategy and raw VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

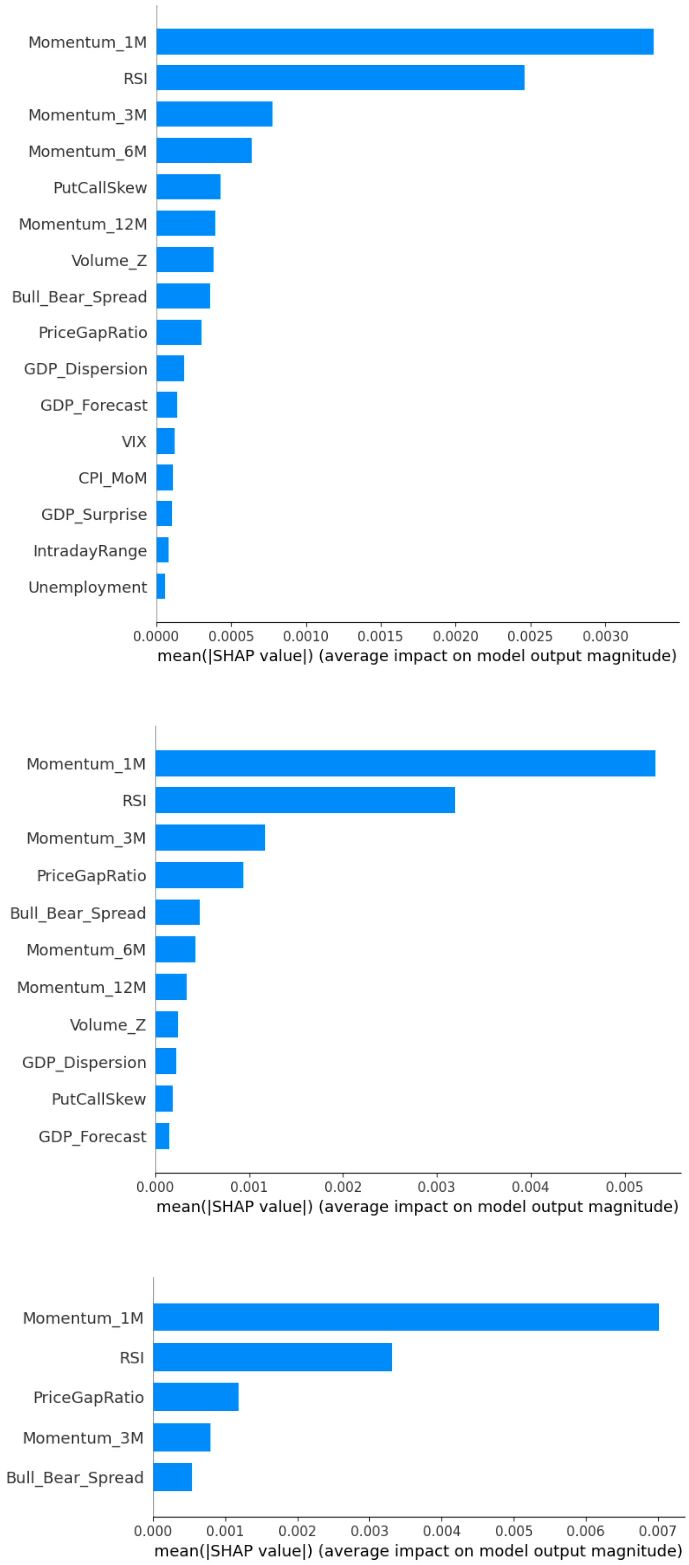

Fig. 4.14 SHAP value comparison for Random Forest using put strategy and raw VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

Excess VRP as Target Variable

In contrast, the Random Forest model was most effective when targeting excess VRP using the put strategy. Model fit improved steadily as features were pruned: the R^2 progressed from -0.14 (all features) to 0.49 (top 10) and finally to 0.69 (top 5), with corresponding RMSE values decreasing from 0.0201 to 0.0150.

ML-driven trading results also improved markedly. The full-feature model produced a 7.29% annualized return, which climbed to 9.55% and 10.51% across the top 10 and top 5 subsets, respectively. Sharpe ratios similarly increased to 0.18 in the top 5 feature configuration, higher than any configuration under the straddle strategy. Moreover, ML win rates exceeded 70%, and SPX correlation remained strong (above 0.80), indicating that the model effectively captured risk-on/risk-off sentiment across market regimes.

Figures 4.15 and 4.16 visually reinforce this result. NAV trajectories indicate smooth and sustained outperformance when the model relies on compact, relevant feature sets. SHAP plots highlight the recurring dominance of VIX, short-term momentum, and macro signals (especially GDP Dispersion and CPI_MoM), which likely served as proxies for behavioral overreaction and volatility dislocation.

Random Forest models performed more robustly under the put-only strategy and when predicting excess VRP, particularly with smaller, behaviorally relevant feature subsets. While the straddle strategy remained consistent in terms of raw returns, regardless of model predictions, the put strategy showed a strong alignment between statistical fit and market performance, especially when feature dimensionality was reduced. This relationship is further discussed in Section 4.5 in the context of comparative model analysis and time-period sensitivity.

doi:10.6342/NTU202503618

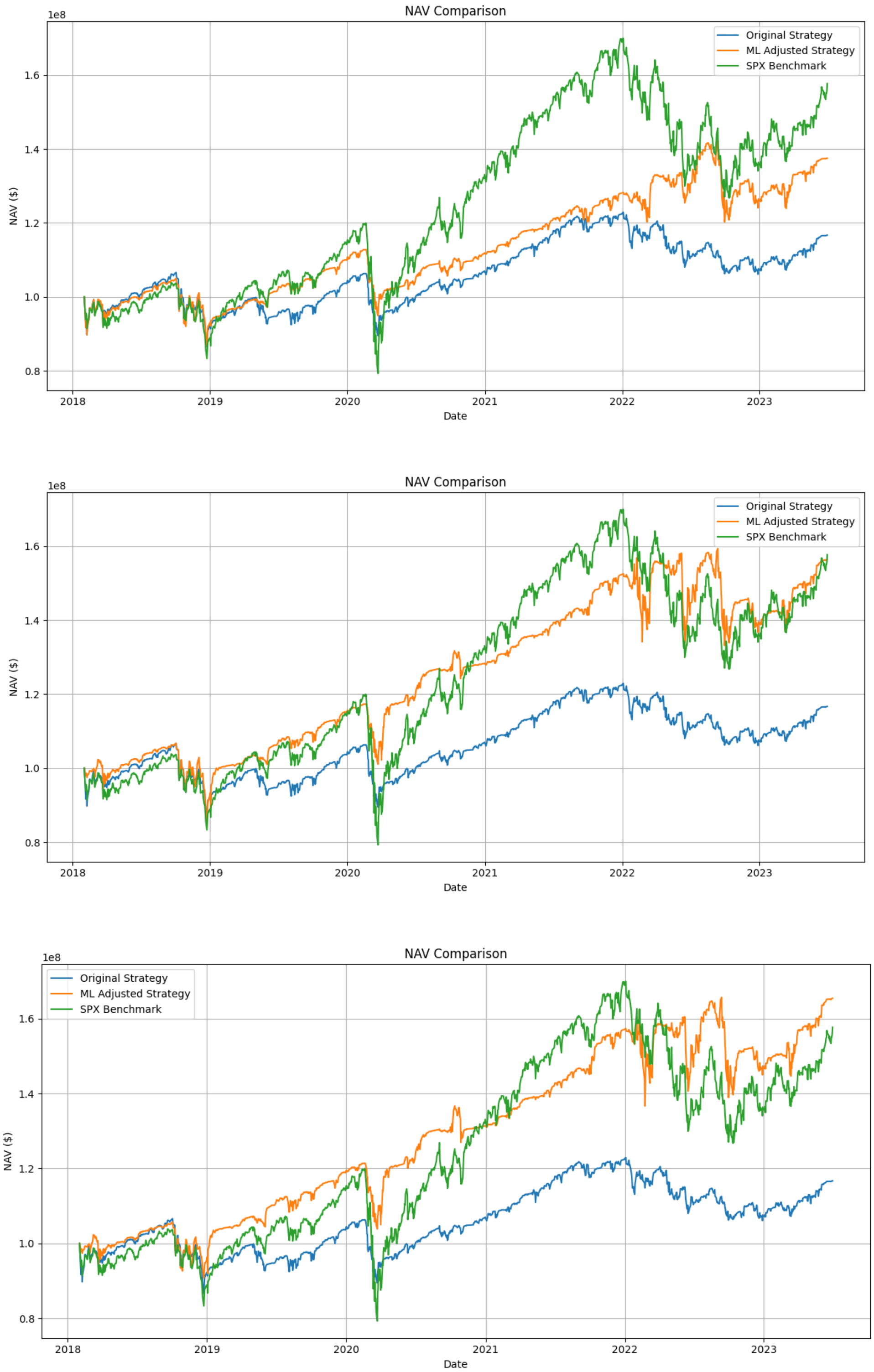

Fig. 4.15 Performance (NAV) comparison for Random Forest using put strategy and excess VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

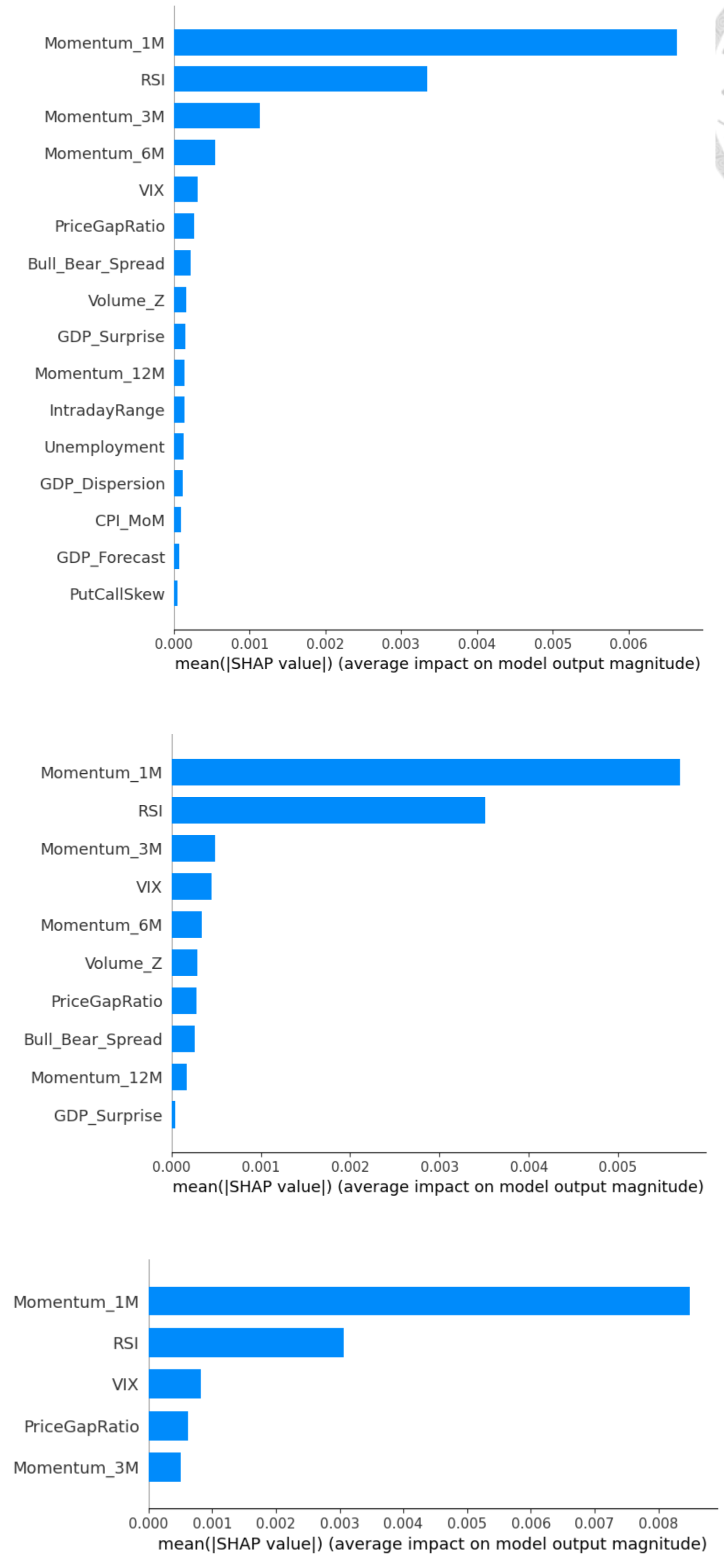

Fig. 4.16 SHAP value comparison for Random Forest using put strategy and excess VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

4.4 Gradient Boosting Results

4.4.1 Straddle Strategy

VRP as Target Variable

When trained on the straddle strategy using raw VRP as the target variable, the Gradient Boosting model showed only marginal improvements over the Random Forest counterpart. With all features included, the model achieved an R^2 of 0.31 and an RMSE of 0.0267. Narrowing the feature space modestly improved the model fit: the top 10 features yielded an R^2 of 0.37 and an RMSE of 0.0258, while the top 5 features showed a slight regression with an R^2 of 0.32 and an RMSE of 0.0264.

Despite the better fit relative to Random Forest, the ML-driven portfolios failed to enhance trading outcomes. All feature subsets resulted in flat to negative annualized returns (All: -0.11%, Top 10: -0.17%, Top 5: -0.55%) and negative Sharpe ratios, while the baseline straddle strategy (with no dynamic scaling) yielded a stable annualized return of 0.83% with a Sharpe of -0.05. NAV comparisons in Figure 4.17 and SHAP visualizations in Figure 4.18 further confirm this underperformance: even though interpretability remained high and feature importance rankings were consistent across models, the extracted signals did not translate into actionable predictive value for VRP timing.

SHAP analysis consistently identified Momentum_1M, RSI, GDP forecasts, and volatility proxies, such as VIX and PutCallSkew, as the top contributors. Their persistent appearance suggests they hold latent explanatory power for the straddle's implied volatility dynamics, even if the model struggled to convert this insight into profit.

doi:10.6342/NTU202503618

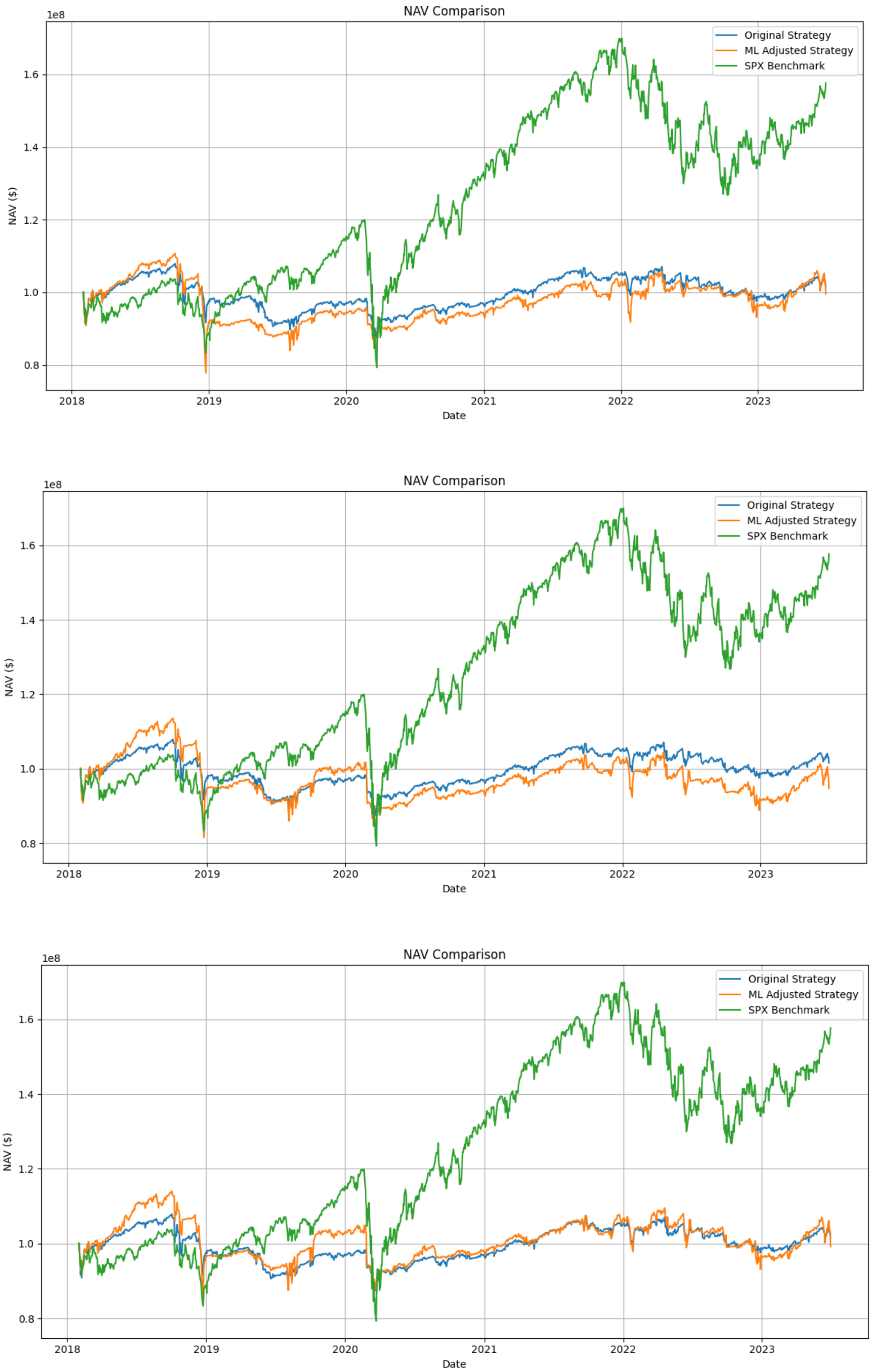

Fig. 4.17 Performance (NAV) comparison for Gradient Boosting using straddle strategy and raw VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

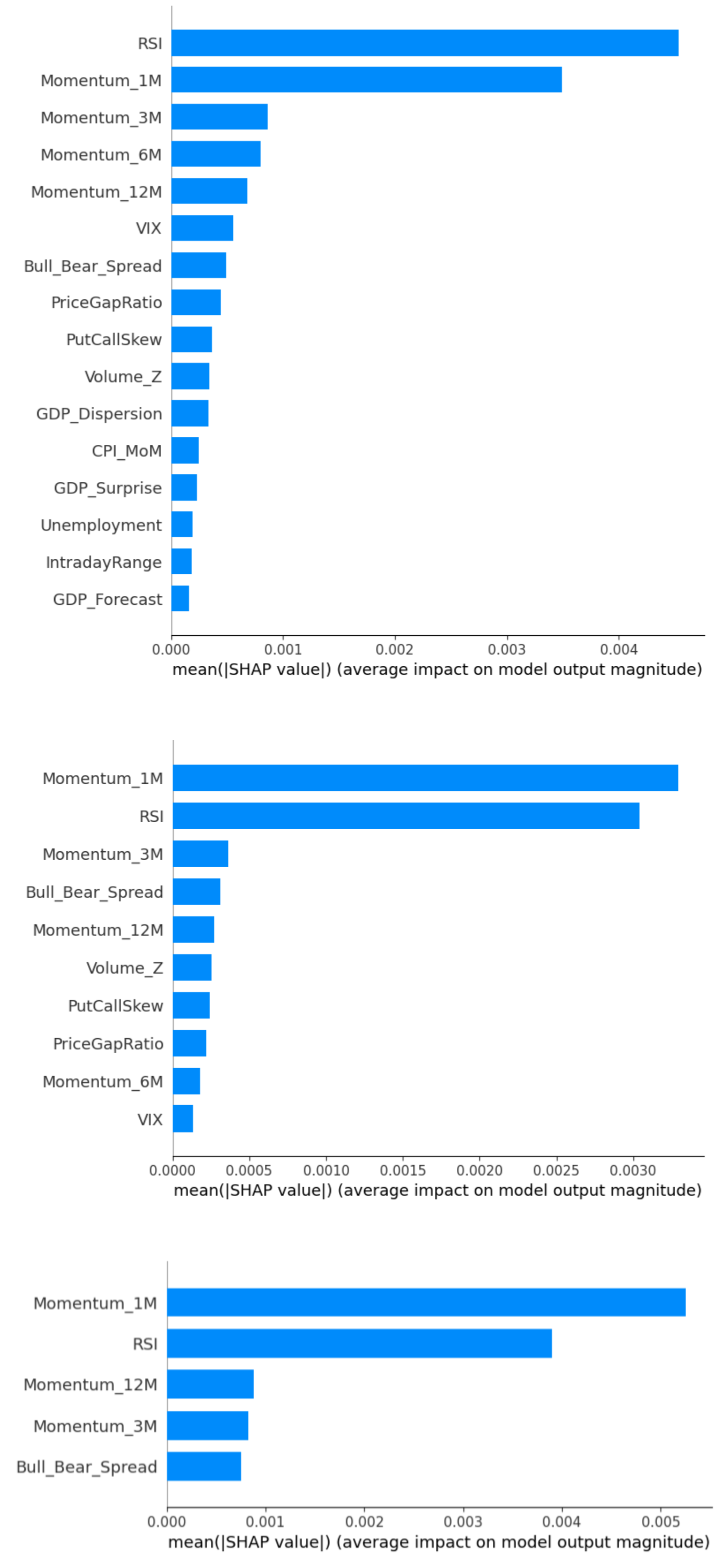

Fig. 4.18 SHAP value comparison for Gradient Boosting using straddle strategy and raw VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

Excess VRP as Target Variable

Performance improved substantially when the target variable shifted to excess VRP. Using all features, the GB model achieved an R^2 of 0.42 and an RMSE of 0.0254. The top 10 feature model improved to an R^2 of 0.48 (RMSE: 0.0234), while the top 5 model peaked at an R^2 of 0.51 and RMSE of 0.0236.

However, these improvements in statistical metrics were again not reflected in trading performance. ML portfolios returned -0.10% (All), -0.51% (Top 10), and -1.04% (Top 5) annually, underperforming the static straddle strategy, which remained flat at 0.83%. As shown in Figure 4.19, the NAV curve initially showed promise. However, it flattened or declined in the latter half of the sample period (2022–2023), possibly indicating structural changes in VRP behavior or a loss of model generalization during volatile macroeconomic regimes.

Despite this, SHAP value analyses again emphasized consistent behavioral and macroeconomic signals. Momentum_1M, RSI, GDP_Surprise, CPI_MoM, and VIX featured prominently, suggesting that excess VRP predictability may be more aligned with relative sentiment and macro shifts than with outright volatility levels.

doi:10.6342/NTU202503618

Fig. 4.19 Performance (NAV) comparison for Gradient Boosting using straddle strategy and excess VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

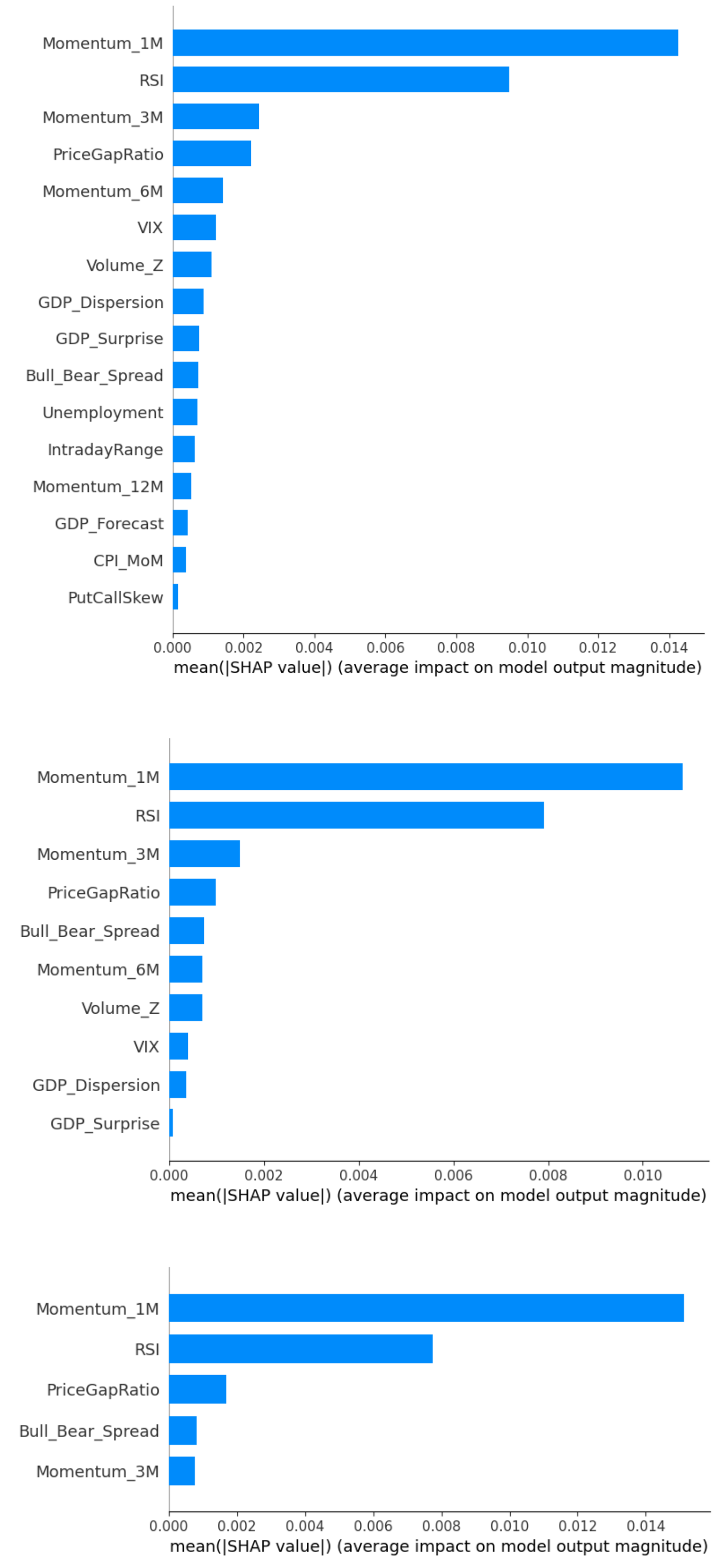

Fig. 4.20 SHAP value comparison for Gradient Boosting using straddle strategy and excess VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

4.4.2 Put Strategy

VRP as Target Variable

The Gradient Boosting model's application to the put strategy using raw VRP as the target produced moderate success in statistical terms. With all features, the model achieved an R^2 of 0.43 and an RMSE of 0.0202. Limiting to the top 10 features improved the R^2 to 0.58, and the top 5 features achieved the highest R^2 of 0.70 and the lowest RMSE of 0.0145.

However, the ML-predicted portfolio struggled to translate statistical power into trading gains. Returns remained flat or negative across all feature sets: All (0.99%), Top 10 (0.29%), and Top 5 (-0.05%), compared to a consistent base strategy return of 3.99%. Sharpe ratios were weak or negative, and drawdowns were materially higher for the ML portfolio, especially when all features were used.

Figures 4.21 and 4.22 confirm these findings. The most influential predictors based on SHAP values included technical and sentiment-based indicators such as RSI, Momentum_1M, PutCallSkew, and VIX, highlighting that investor positioning and volatility skew were more predictive of raw VRP than macroeconomic surprises. Traditional macroeconomic indicators, such as GDP Surprise, Unemployment, and CPI MoM, ranked lower in importance, suggesting that raw VRP may be more responsive to market microstructure and behavioral flows than to top-down economic data. However, despite strong model fit metrics, this signal strength did not translate into reliable portfolio performance, possibly due to regime shifts or overfitting to historical patterns.

doi:10.6342/NTU202503618

Fig. 4.21 Performance (NAV) comparison for Gradient Boosting using put strategy and raw VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

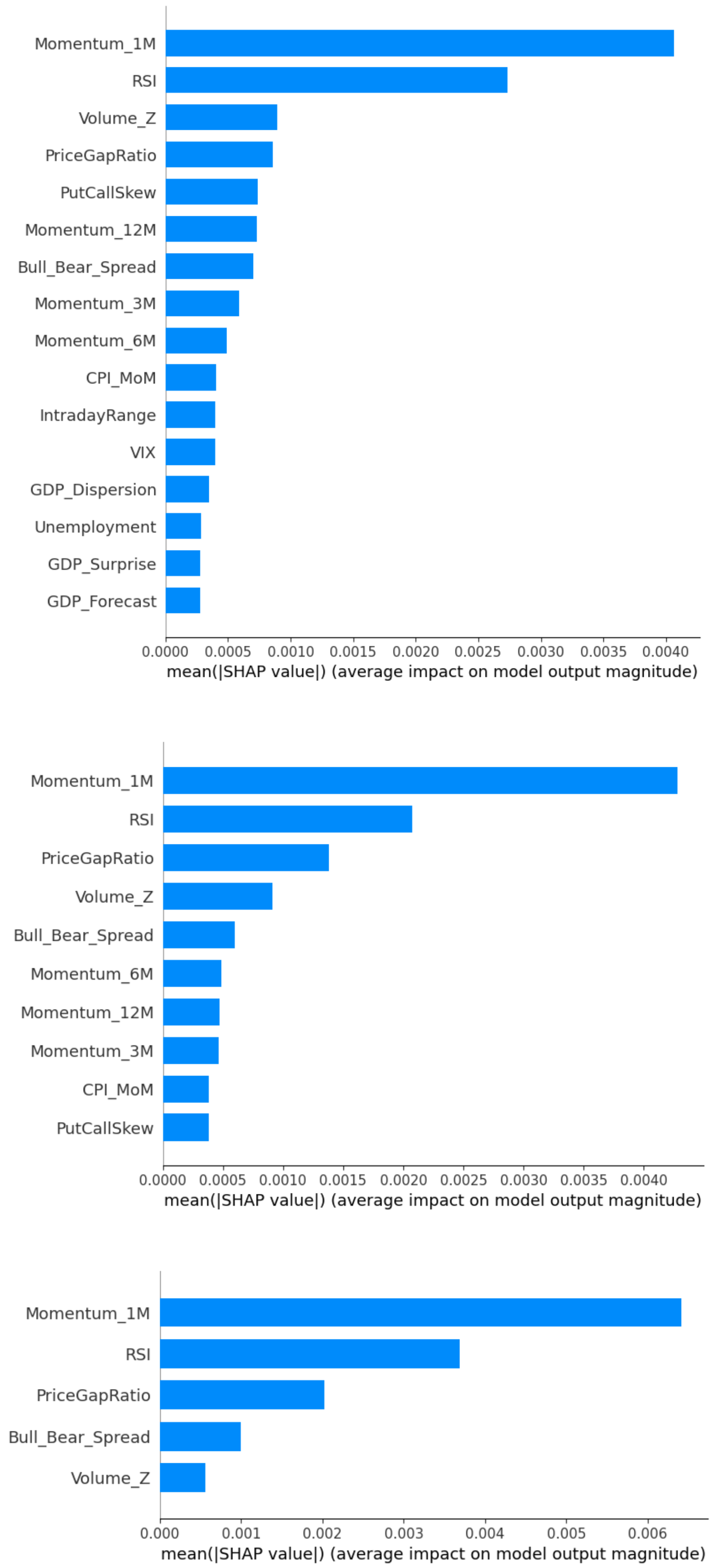

Fig. 4.22 SHAP value comparison for Gradient Boosting using put strategy and raw VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

Excess VRP as Target Variable

The GB model achieved its strongest overall results when applied to the put strategy with excess VRP as the target. The R^2 steadily improved from 0.60 (All) to 0.76 (Top 10) and peaked at 0.75 (Top 5). RMSE correspondingly declined from 0.0181 to 0.0164 and 0.0161. These values represent the best overall model fits in the entire study.

Importantly, economic performance improved in parallel. The ML portfolio returned 7.38% (All), 9.76% (Top 10), and 10.22% (Top 5) annually, significantly outperforming the baseline strategy (3.99%) and rivaling the SPX benchmark (10.97%). The Sharpe ratio reached 0.17 under the top 5 feature set, with the lowest drawdown (-0.19%) and highest win rate (70.31%).

Figures 4.23 and 4.24 illustrate the strong consistency of NAV and SHAP across feature subsets. The top features—VIX, RSI, Momentum_1M, GDP Dispersion, and CPI_MoM—highlight the critical role of sentiment, macro volatility, and behavioral disequilibrium in capturing the excess volatility premium. Notably, the reduced feature sets outperformed the full model both statistically and economically, reinforcing the importance of model parsimony in high-dimensional macro-financial contexts.

In summary, Gradient Boosting models showed mixed results. They underperformed in the straddle strategy regardless of the target variable, likely due to lower signal strength in symmetric volatility exposures. However, in the put strategy, especially when targeting excess VRP, they produced the highest predictive accuracy and strongest economic results in the entire analysis. These findings underscore the value of asymmetric VRP strategies and disciplined feature selection in constructing machine-learning-driven volatility portfolios.

doi:10.6342/NTU202503618

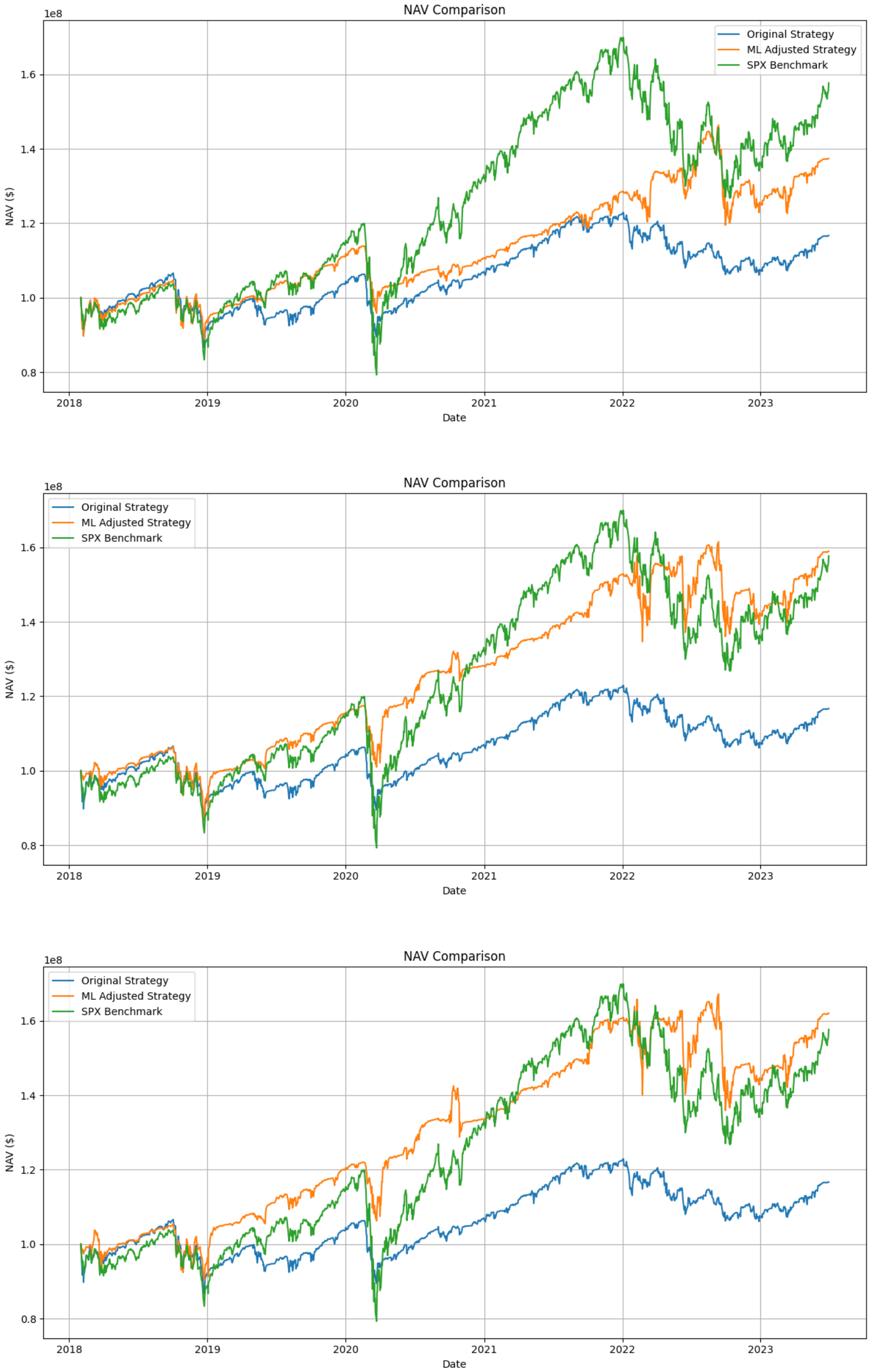

Fig. 4.23 Performance (NAV) comparison for Gradient Boosting using put strategy and excess VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

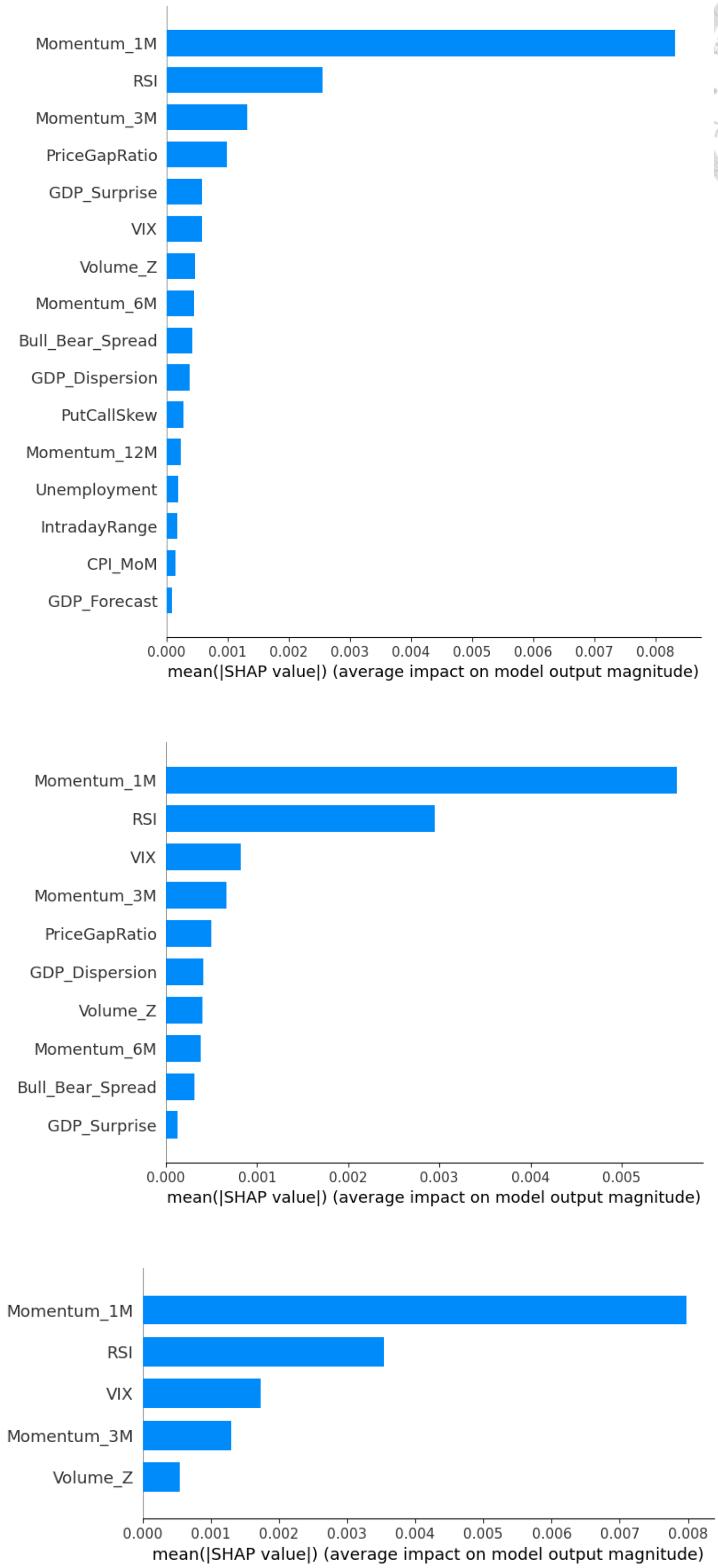

Fig. 4.24 SHAP value comparison for Gradient Boosting using put strategy and excess VRP as the target with all features (top), top 10 features (middle), top 5 features (bottom)

doi:10.6342/NTU202503618

4.5 Model Comparison and Interpretation

This section synthesizes the empirical findings across the linear and tree-based models—Ridge Regression, Random Forest, and Gradient Boosting—to evaluate their relative predictive accuracy, portfolio performance, feature behavior, and practical implications across the four strategy-target configurations: Straddle with Raw VRP, Straddle with Excess VRP, Put with Raw VRP, and Put with Excess VRP. The analysis draws on Tables 4.2 and 4.3, as well as Figures 4.9 through 4.16, which include performance metrics, NAV plots, SHAP visualizations, and feature importances.

4.5.1 Predictive Accuracy

From a statistical modeling standpoint, Gradient Boosting consistently achieved the highest in-sample R^2 across nearly all configurations when trained on SHAP-selected top 5 or top 10 features. For instance, the Gradient Boosting Put strategy with raw VRP as the target attained an R^2 of 0.8299, the highest among all model-strategy combinations. Similarly, the Gradient Boosting Straddle strategy with excess VRP recorded an R^2 of **0.7938** with the top 5 features. In contrast, Ridge Regression models displayed modest predictive power (e.g., $R^2 = 0.2943$ for Put with excess VRP), while Random Forest models exhibited more variability, performing well with fewer features but prone to overfitting when trained on all features (e.g., negative R^2 for Straddle with raw VRP).

Interestingly, the most predictive target variable also varied by strategy. For the Put strategy, models trained to predict raw VRP generally achieved higher R^2 than those using excess VRP, particularly in tree-based models. In contrast, Straddle strategies appeared to benefit more from predicting excess VRP, likely due to the straddle's symmetrical exposure and sensitivity to equity beta that excess VRP helps to neutralize.

doi:10.6342/NTU202503618

4.5.2 Market Performance

Despite the superior statistical fit of tree-based models, portfolio-level performance tells a more nuanced story. When examining strategy returns, Sharpe ratios, drawdowns, and correlations, the Put strategy outperforms Straddle strategies across all models and target variables. For instance, (1) The Ridge Regression Put strategy with excess VRP achieved an annualized return of 14.98% and a Sharpe ratio of 0.29, notably higher than any Ridge-based straddle configuration. (2) Tree-based Put strategies also delivered strong results, with Gradient Boosting (raw VRP) reaching up to 10.22% annualized return and a Sharpe ratio of 0.17. However, Sharpe ratios plateaued relative to Ridge in some variants.

The Straddle strategy, by contrast, exhibited flat or negative Sharpe ratios and negligible alpha, particularly for models targeting raw VRP. For instance, both Random Forest and Gradient Boosting models exhibited strategy Sharpe ratios of -0.05 when applied to the Straddle strategy, regardless of the predictive R^2 values. This suggests that predictive accuracy alone is insufficient to ensure superior market performance, especially when the strategy construction involves both long and short delta exposures and hence interacts more complexly with directional market trends.

The Put strategy's consistent outperformance may be attributed to the underlying macroeconomic conditions during the sample period (2018–2023). This period was predominantly characterized by a bull market, especially post-2020, during which put options consistently decayed in value due to low realized volatility and positive equity returns. By targeting VRP, particularly when predicting when puts are overpriced, the strategy likely captured volatility risk premium more effectively than the straddle, which requires accurate timing of both tails.

doi:10.6342/NTU202503618

4.5.3 Unexpected Strength of Ridge Regression for Put Strategies

A notable empirical result is the unexpected competitiveness of Ridge Regression when applied to the Put strategy. Despite its simplicity and lack of non-linear modeling, Ridge achieved performance metrics comparable to or even exceeding tree-based alternatives in some scenarios. This was the case regardless of the target variable, with Ridge regression outperforming the tree models for both raw and excess VRP target variables.

There are several potential explanations for this phenomenon. First, overfitting is a documented issue in high-dimensional tree models, particularly in relatively small datasets with time series structure. Ridge Regression, with its penalization of large coefficients and linear bias, may have provided better out-of-sample generalization, especially when signals were driven by a few dominant behavioral indicators (e.g., RSI, PutCallSkew, Momentum_1M). Second, linear relationships may suffice to capture VRP timing in bull market regimes, where behavioral sentiment and momentum tend to align in predictable ways. Finally, feature engineering may have disproportionately benefited linear models. Features like RSI, PutCallSkew, and VIX are already nonlinear transformations of prices or returns, potentially diminishing the marginal benefit of additional model complexity in capturing their effects.

However, even with this relative strength, Ridge Regression-based Put strategies underperformed the S&P 500 throughout 2021. This was possibly due to the extraordinary post-pandemic bull run, where equity markets delivered outsized gains while implied volatility steadily declined. In such environments, delta-hedged short put strategies generate limited upside because they are designed to harvest volatility mispricing rather than directional equity returns. Moreover, the persistent collapse in VIX during 2021 reduced the magnitude of the volatility risk premium, resulting in

doi:10.6342/NTU202503618

thinner option selling profits. The Ridge model, while able to capture periods of stable VRP, could not fully compensate for the structural underperformance against a strongly trending equity index. Behavioral signals like RSI and momentum also became less predictive in 2021, as market returns were driven more by macroeconomic tailwinds (e.g., fiscal stimulus, reopening trades) than by volatility risk aversion, leading to muted ML-driven scaling and lower NAV growth relative to SPX.

4.5.4 Feature Interpretability and Strategy Alignment

The SHAP analysis across Random Forest and Gradient Boosting models confirms that feature importance rankings are relatively consistent across strategies. For Put strategies, dominant predictors include: (1) RSI – often the most important feature across all configurations, (2) PutCallSkew – a measure of downside fear. (3) Momentum_1M – short-term price direction, and (4) VIX and Volume_Z – proxies for volatility expectations and abnormal trading activity.

In contrast, macro features such as GDP Surprise, Unemployment, and CPI_MoM ranked lower in most configurations, particularly when predicting raw VRP. This suggests that behavioral and sentiment-driven features are more effective in predicting volatility mispricing than traditional macroeconomic data.

Interestingly, feature rankings did not always align with return-generating power. In several cases, models with better in-sample fit (e.g., top 10 or top 5 features) delivered worse out-of-sample returns, particularly in Straddle strategies. This reflects a well-known tradeoff in financial modeling: excessive optimization of predictive metrics often fails to generalize under regime shifts, especially for strategies sensitive to skew and kurtosis.

doi:10.6342/NTU202503618

4.5.5 Practical Takeaways

From a practical perspective, the results highlight several key insights:

1. Strategy selection dominates model complexity. The choice between Put and Straddle strategies had a larger impact on return outcomes than the choice of model. Even simple models, such as Ridge, outperformed complex ones when aligned with market trends and option payoff structures.
2. Model complexity improves the statistical fit but does not necessarily return the best results. Gradient Boosting achieved the best predictive accuracy but often underperformed Ridge in real-world metrics due to overfitting or market shifts.
3. Behavioral features are critical. RSI, Momentum, and PutCallSkew consistently ranked as the most important features across all model families and strategies, reaffirming the importance of behavioral finance in volatility pricing.
4. Market environment sensitivity is significant. Tree-based models performed better from 2018 to 2022 but struggled afterward, as indicated by NAV plots and declining Sharpe ratios. This suggests that changes in market structure (e.g., inflation shock, Fed tightening) may have altered the behavioral dynamics underlying VRP.

doi:10.6342/NTU202503618

5. Conclusion

5.1 Summary of Findings

This thesis explores the predictive modeling of VRP profitability using machine learning, with a focus on two delta-hedged options strategies: the short straddle and the short put. Three model classes—Ridge Regression, Random Forest, and Gradient Boosting—are evaluated using two target variables: raw VRP and excess VRP (VRP net of contemporaneous S&P 500 returns). This design enables a nuanced examination of how well models capture volatility-related alpha compared to simple equity beta.

The results yield several important findings. First, non-linear models (particularly Random Forest) consistently outperform linear Ridge Regression in terms of predictive accuracy (as measured by R^2 and RMSE). While Ridge Regression offers better interpretability and more stable coefficients, its ability to capture non-linear structures in VRP profitability is limited, especially in the context of the straddle strategy.

Second, contrary to expectations, Random Forest generally outperforms Gradient Boosting, especially in the raw VRP configurations and in predicting put strategy profitability. While Gradient Boosting produces more compact trees and sometimes benefits from better generalization, its out-of-sample economic performance often lags that of Random Forest, suggesting that the ensemble variance control and robustness of Random Forest make it better suited for the noisy financial environment of volatility trading.

A particularly notable finding is the consistent superiority of put-based strategies over straddle strategies across all model classes and configurations. Put strategies

doi:10.6342/NTU202503618

exhibit significantly higher annualized returns (ranging from 10% to over 14% in the best cases), higher Sharpe ratios, and lower drawdowns, especially when modeled with Random Forest using the excess VRP target. This aligns with behavioral theories positing that downside insurance carries a persistent premium due to investor aversion to tail risks.

Interestingly, Ridge Regression's put strategy, which utilizes excess VRP, while exhibiting a low R^2 and high RMSE, delivers the strongest realized economic performance, suggesting that a weak statistical fit may still yield high-quality signals under certain market conditions. This discrepancy raises important questions about the relationship between predictive accuracy and economic utility in financial modeling.

Finally, the choice of target variable proves critical. For straddle strategies, excess VRP significantly enhances model performance by stripping away market directionality. For put strategies, however, raw VRP remains effective in terms of predictive power—likely because put premiums already embed downside beta, rendering excess VRP transformations less additive.

SHAP-based feature selection further enhances performance. Reducing the feature set to the top 10 or 5 most impactful variables not only simplifies the model but often improves out-of-sample generalization. Key features across models include VIX, momentum indicators, RSI, and macroeconomic surprises such as GDP forecasts and CPI, which frequently rank among the most predictive inputs.

5.2 Practical Implications

The findings from this research provide actionable insights for practitioners engaged in quantitative volatility trading. First, the put-only strategy, particularly when combined with excess VRP targeting, appears to offer a robust, high-return opportunity

doi:10.6342/NTU202503618

with manageable risk. This makes it especially appealing for portfolio overlays or systematic volatility harvesting strategies.

Second, the results reinforce the value of ensemble-based machine learning models, which handle noisy and non-stationary financial data more effectively than linear alternatives. However, model selection must go beyond predictive accuracy. As demonstrated, the most statistically accurate models (e.g., Gradient Boosting) do not consistently deliver the best trading results, and linear models with relatively poor R^2 values may still yield alpha.

The superior performance of simpler models in certain market environments also underscores the importance of model simplicity, robustness, and interpretability, especially in institutional contexts with regulatory and fiduciary oversight. Tools like SHAP values facilitate model transparency, aiding in feature selection and enhancing trust and communication between data scientists, portfolio managers, and stakeholders.

Finally, the temporal breakdown of performance reveals important regime dependency. Tree-based models performed well from 2018 to 2021, a period marked by macroeconomic stability and strong risk appetite, but degraded significantly during 2022–2023, which was characterized by inflation, rate hikes, and geopolitical volatility. This suggests that real-time monitoring, adaptive retraining, and potential integration of regime-switching frameworks are critical for maintaining performance over time.

5.3 Future Research Directions

This thesis opens several promising avenues for further investigation. A key direction is to ensure model robustness across different market environments. The degradation in model efficacy post-2022 suggests that volatility-related alpha is not time-invariant. Future work could explore regime-aware models, such as hidden

doi:10.6342/NTU202503618

Markov models or conditional tree ensembles, to adapt model behavior to shifting macroeconomic environments.

Another direction is the integration of alternative data and architectures. Although tree-based models excel at handling structured data, incorporating unstructured sources, such as news sentiment, option order flows, or social media, using deep learning (e.g., LSTMs or transformers) may yield further improvements in predictive capacity and adaptability.

Feature engineering can also be expanded. While this study focused on macroeconomic, behavioral, and technical indicators, intraday volatility signals, options surface metrics, and skewness/kurtosis measures may enhance model resolution and allow finer-grained signal calibration.

Finally, transaction costs and market frictions remain crucial considerations. While this research assumes frictionless execution, future work should account for bid-ask spreads, slippage, and capital constraints in simulating real-world implementations. This is particularly important for put strategies, where liquidity may dry up during stress events, potentially amplifying risk or distorting model signals.

5.4 Concluding Remarks

This thesis contributes to the intersection of quantitative finance and machine learning by empirically validating that VRP profitability is both forecastable and tradable using systematic models. By comparing linear and non-linear architectures across two distinct strategies and target variable frameworks, this research presents a comprehensive view of how volatility premia can be extracted from market data.

Despite the success of tree-based models, the findings caution against over-reliance on any single architecture. Model performance is context-sensitive, and a high

doi:10.6342/NTU202503618

in-sample R^2 does not guarantee out-of-sample profitability. As such, the successful implementation of ML-driven volatility strategies requires not only technical modeling skills but also macroeconomic awareness, robust risk management, and ongoing adaptation to evolving financial regimes.

In sum, this study affirms that machine learning, when carefully constructed, validated, and interpreted, can serve as a powerful tool for uncovering alpha in the volatility space. However, like all tools in finance, its edge lies in how it is used, updated, and interpreted, rather than its complexity alone.

doi:10.6342/NTU202503618

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2003). Micro effects of macro announcements: Real-Time price discovery in foreign exchange. *American Economic Review*, *93*(1), 38–62. https://doi.org/10.1257/000282803321455151

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *The Journal of Economic Perspectives*, *21*(2), 129–151. https://doi.org/10.1257/jep.21.2.129

Bali, T. G., Beckmeyer, H., Moerke, M., & Weigert, F. (2021). Option Return Predictability with Machine Learning and Big Data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3895984

Bali, T. G., Brown, S. J., & Tang, Y. (2016). Is economic uncertainty priced in the Cross-Section of individual stocks? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2812967

Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, *94*(1–2), 181–238. https://doi.org/10.1016/s0304-4076(99)00021-4

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*. https://doi.org/10.1142/9789814759588_0001

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623–685. https://doi.org/10.3982/ecta6248

Bollen, N. P. B., & Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *The Journal of Finance*. https://doi.org/10.1111/j.1540-6261.2004.00647.x

doi:10.6342/NTU202503618

Bollerslev, T., Gibson, M. S., & Zhou, H. (2011). Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities. *Journal of Econometrics.* https://doi.org/10.1016/j.jeconom.2010.03.033

Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *SSRN Electronic Journal*. https://doi.org/10.1093/rfs/hhp008

Carr, P., & Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, *22*(3), 1311–1341. https://doi.org/10.1093/rfs/hhn038

Chordia, T., Subrahmanyam, A., & Roll, R. (2000). Market liquidity and trading activity. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.237674

Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*. https://doi.org/10.1016/S0304-405X(98)00034-8

Christensen, K., Siggaard, M., & Veliyev, B. (2022). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, *21*(5), 1680–1727. https://doi.org/10.1093/jjfinec/nbac020

Corradi, V., Distaso, W., & Mele, A. (2012). Macroeconomic Determinants of stock market volatility and volatility Risk-Premiums. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2005021

Drechsler, I., & Yaron, A. (2011). What's Vol got to do with it. *The Review of Financial Studies, 24(1), 1–45. http://www.jstor.org/stable/40985815*

French, K., Schwert, G., & Stambaugh, R. (1987). Expected stock returns and volatility. *Journal of Financial Economics*. https://doi.org/10.1016/0304-405X(87)90026-2

doi:10.6342/NTU202503618

Garman, M. B., & Klass, M. J. (1980). On the Estimation of Security Price Volatilities from Historical Data. *The Journal of Business*, *53*(1), 67–78. http://www.jstor.org/stable/2352358

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Guo, I., & Loeper, G. (2020). The Volatility Risk Premium: An Empirical Study on the S&P 500 Index. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3739933

GuyonIsabelle, & ElisseeffAndré. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*. https://doi.org/10.5555/944919.944968

Jackwerth, J. C., & Rubinstein, M. (1996). Recovering Probability Distributions from Option Prices. *The Journal of Finance*, *51*(5), 1611. https://doi.org/10.2307/2329531

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. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Londono, J. M., Ma, S., & Wilson, B. A. (2025). *Costs of rising uncertainty*. https://www.federalreserve.gov/econres/notes/feds-notes/costs-of-rising-uncertainty-20250424.html

Lou, D., Polk, C., & Skouras, S. (2019). A tug of war: Overnight versus intraday Expected Returns. *Journal of Financial Economics.* https://doi.org/10.1016/j.jfineco.2019.03.011

Lundberg, S., & Lee, S. (2017, May 22). A unified approach to interpreting model predictions. *arXiv.org*. https://arxiv.org/abs/1705.07874

doi:10.6342/NTU202503618

Merton, R. C. (1975). *Option pricing when underlying stock returns are discontinuous*. https://dspace.mit.edu/handle/1721.1/1899

Rudin, C. (2018). *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*. arXiv.org. https://arxiv.org/abs/1811.10154

Smales, L. (2014). News sentiment and the investor fear gauge. *Finance Research Letters*. https://doi.org/10.1016/j.frl.2013.07.003

Tang, A. (2023). *Option trading strategies to harvest the volatility risk premium*. http://arks.princeton.edu/ark:/88435/dsp01zw12z858k

doi:10.6342/NTU202503618