

Exploring Leverage, Sentiment and Inversion Effects on the Dynamics of Implied Volatility

Zifeng Tang^{1,2}, Yunfei Song³ and André Jacob^{1,2}

¹*Department of Mathematics, ETH Zürich;* ²*Department of Banking and Finance, University of Zürich;* ³*Department of Chemistry, University of Zürich*

Abstract. In this study, we examine the impact of market returns and sentiment on the dynamics of implied volatility of S&P 500 options using XGBoost and LightGBM models. These models demonstrate enhanced efficiency and accuracy, outperforming traditional neural network-based methods for this specific issue. Our further analysis confirms the significant influence of leverage, sentiment, and inversion effects on the movement of implied volatility. Furthermore, we introduce the HWRTI(n_1, n_2, V_0, r_0) econometric model, which refines existing analytical approaches and offers clearer insights into the relationships between market returns, sentiment, and implied volatility for different S&P 500 options. The adaptability of this model provides novel insights into the dynamics of implied volatility, proving beneficial for advancing both financial theory and practical application.

Keywords: Leverage Effect, Sentiment Effect, Inversion Effect, Implied Volatility Movement

JEL Classification: C58, G13

1. Introduction

The study of implied volatility change stands at the forefront of contemporary financial analysis. These changes not only encapsulate the market's evolving expectations of future volatility but also serve as a barometer for market sentiment, liquidity, and macroeconomic shifts. Understanding the nuances of implied volatility is paramount for developing more responsive and resilient financial models. This study en-

deavors to unravel the intricate dynamics of implied volatility changes, offering valuable insights for investors, portfolio managers, and market theorists in their pursuit of more informed and strategic decision-making in an ever-fluctuating financial landscape.

Building on this understanding of implied volatility, it's crucial to delve into specific factors that influence it. One such factor is the leverage effect, a fundamental concept in financial theory that interplays significantly with market sentiment and returns, as our study will demonstrate. A pivotal topic in the financial study, traces its origins to Black [1976], which highlighted the negative relationship between equity prices and volatility. Following Black, Christie [1982], leveraging Modigliani and Miller's theories on company financial leverage, provided a detailed explanation of this effect, linking it to the concept of leverage. However, subsequent study by French et al. [1987], Campbell and Hentschel [1992], Bekaert and Wu [2000], Bollerslev et al. [2006], Hens and Steude [2009], and Hasanhodzic and Lo [2019] broadened this understanding, presenting new evidence and perspectives, suggesting a reverse causality between the changes in volatility and equity price. Cont and Da Fonseca [2002] and Poulsen et al. [2009], have documented that this effect extends to implied volatilities and underlying returns.

While the qualitative aspects of the leverage effect are well-established, the field still lacks a robust quantitative analysis of the dynamics between underlying returns and implied volatilities. In this realm, the work by Hull and White [2017] on optimizing delta hedging, particularly for the S&P 500 options, indirectly contributes to this understanding. Their model, fundamentally based on linear regression analysis, while focused on delta hedging, provides insights into the implied volatility surface's response to underlying asset price changes. Cao et al. [2020] embraced this innovative approach but ventured further into the realm of machine learning. They harness the power of neural networks that by the Universal Approximation Theorem can approximate any continuous function arbitrarily well and consider the expected implied volatility change as a function of underlying asset price change, moneyness and time to maturity to gain a 10.72% improvement. They also have documented a further improvement of 62.12% considering the VIX index as an additional variable the function depends on.

Expanding on this groundwork, we apply XGBoost and LightGBM, two advanced but classic machine learning methodologies, to investi-

gate changes in implied volatility. Both models are adept at capturing non-linear patterns in data, a characteristic they share with neural networks. XGBoost constructed by Chen and Guestrin [2016] is noted for its effectiveness in various machine learning applications, owing to its sophisticated handling of diverse datasets and its ability to balance bias and variance. Subsequently, Ke et al. [2017] introduced LightGBM, notable for its pioneering techniques including Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). These methods significantly enhance the handling of large and high-dimensional datasets, making them well-suited candidates for addressing the issues presented in this study.

Having established the theoretical underpinnings of our approach with XGBoost and LightGBM, we now turn to the practical aspects of our study, detailing the model development, data handling, and analysis in the subsequent sections. Section 2 outlines the concept of implied volatility, elaborates on the construction of our models, and describes the loss measures used for model assessment. In Section 3, we discuss the handling of our dataset, the training process for various models, and conduct a comparative analysis of their performances. Section 4 offers a comprehensive analysis of the impact of different features on implied volatility movements and identifies three key effects. Section 5 extends this analysis by employing an adaptive analytical model to validate and clarify the effects uncovered in Section 4. Section 6 consolidates our insights into the dynamics of implied volatility and offers recommendations for further investigation in this field.

2. Methodology

This section details the models employed to capture the movements of implied volatility, along with the metrics used to evaluate their predictive accuracy.

2.1 Implied volatility and Greek

We assume that an underlying asset satisfies Geometric Brownian Motion (GBM)

$$dS_t = \mu S_t dt + \sigma S_t dW_t \quad (1)$$

where W_t is Wiener Brownian Motion under physical probability measure \mathbb{P} . μ is a constant drift and σ is a constant volatility.

If further assume constant dividend q , constant riskless asset return r , and arbitrage-free market, then the European call and put option pricing formula are as follows

$$\begin{aligned} C_t &= S_t e^{-q(T-t)} \Phi(d_1) - K e^{-r(T-t)} \Phi(d_2) \\ P_t &= -S_t e^{-q(T-t)} \Phi(-d_1) + K e^{-r(T-t)} \Phi(-d_2) \\ d_1 &= \frac{\ln(\frac{S_t}{K}) + (r - q + \frac{\sigma^2}{2})(T - t)}{\sigma \sqrt{T - t}} \\ d_2 &= d_1 - \sigma \sqrt{T - t} \end{aligned} \tag{2}$$

$T - t$ is the time to maturity and is often redefined as $T := T - t$ without any gain. It is only kept in this subsection for consistency purposes with Equation (1).

The concept of implied volatility relates is the estimated volatility level of the underlying asset that, when used in the above option pricing formulas, yields a theoretical price that matches the actual market price of the corresponding option, i.e.

$$\begin{aligned} C_t(\sigma_{imp}^{call}) &= C_{t,Market} \\ P_t(\sigma_{imp}^{put}) &= P_{t,Market} \end{aligned} \tag{3}$$

Note that the implied volatilities for call and put options are different even though they have identical strike prices and time to maturity, but we simplify notation using σ_{imp} when the context is clear.

The Greeks are a set of measures that indicate how sensitive the price of a derivative, like an option, is to variations in one or more underlying factors that influence the value of that financial instrument or a portfolio comprising multiple financial instruments. Their mathematical essence is the partial derivative. The Greeks particularly of interest in this study are Delta: δ and Vega: ν . They are given by

$$\begin{aligned} \delta^{call} &= \frac{\partial C_t}{\partial S_t} = e^{-q(T-t)} \Phi(d_1) \\ \delta^{put} &= \delta^{call} - 1 \end{aligned} \tag{4}$$

$$\begin{aligned} \nu^{call} &= \frac{\partial C_t}{\partial \sigma} = S_t e^{-q(T-t)} \Phi(d_1) \sqrt{T - t} \\ \nu^{put} &= \nu^{call} \end{aligned} \tag{5}$$

The call option price is strictly increasing with respect to S_t because δ^{call} is positive, so is it with respect to σ because ν^{call} is also positive.

The values of δ and ν at σ_{imp} are what is called the practitioner Black-Scholes Delta and Vega. We use the notation δ_{BS} and ν_{BS} . Many research studies use δ_{BS} as a measure of moneyness, with the associated ranges detailed below.

Table 1 δ_{BS} Ranges for OTM/ATM/ITM Call and Put Options

Moneyness	OTM	ATM	ITM
Call	(0, 0.5)	0.5	(0.5, 1)
Put	(-0.5, 0)	-0.5	(-1, -0.5)

2.2 Models

We introduce 7 models that are experimented with in this section. We refer to their abbreviations for simplification in the subsequent discussion.

Table 2 Model Summary

Model	Notation
Hull-White analytic model	HW
Neural Network 3-Feature model	NN-3
Neural Network 4-Feature model	NN-4
XGBoost 3-Feature model	XGBoost-3
XGBoost 4-Feature model	XGBoost-4
LightGBM 3-Feature model	LGBM-3
LightGBM 4-Feature model	LGBM-4

2.2.1 Hull-White Model

The benchmark model used in this study is based on the analytical framework proposed by Hull and White [2017]. It can be expressed as

$$\mathbb{E}(\Delta\sigma_{imp}) = \frac{\beta_1 + \beta_2\delta_{BS} + \beta_3\delta_{BS}^2}{\sqrt{T}} r. \quad (6)$$

where the β_j , $j \in \{1, 2, 3\}$, are unknown parameters that need to be estimated, δ_{BS} is the practitioner Black-Scholes delta, T is the time to maturity and r is the return of an asset. The equation is derived through the following steps. First, by applying Taylor's expansion and making certain assumptions, we can obtain the following expression:

$$\delta_{\text{MV}} = \frac{\partial f_{\text{BS}}}{\partial S} + \frac{\partial f_{\text{BS}}}{\partial \sigma_{\text{imp}}} \frac{\partial E(\sigma_{\text{imp}})}{\partial S} = \delta_{\text{BS}} + \nu_{\text{BS}} \frac{\partial E(\sigma_{\text{imp}})}{\partial S}$$

where δ_{MV} is the minimum variance delta, f_{BS} is the Black-Scholes pricing function, and ν_{BS} is the practitioner Black-Scholes vega.

Then, they made use of the following observations:

1. The difference $\delta_{\text{MV}} - \delta_{\text{BS}}$ is not significantly influenced by option maturity and exhibits an approximately quadratic relationship with δ_{BS} .
2. When dividends are small, $\nu_{\text{BS}} \approx S\sqrt{T}G(\delta_{\text{BS}})$.

Based on these observations, they derived the following equation:

$$\delta_{\text{MV}} = \delta_{\text{BS}} + \frac{\nu_{\text{BS}}}{S\sqrt{T}} (a + b\delta_{\text{BS}} + c\delta_{\text{BS}}^2),$$

from which one can easily obtain 6.

2.2.2 Neural Network Model

Two neural network models proposed by Cao et al. [2020] are implemented: NN-3 and NN-4. The following are abstract functions they approximate, respectively:

$$\mathbb{E}[\Delta\sigma_{\text{imp}}] = f(r, \delta_{\text{BS}}, T) \quad (7)$$

$$\mathbb{E}[\Delta\sigma_{\text{imp}}] = g(r, \delta_{\text{BS}}, T, V) \quad (8)$$

where r is the return of an asset, δ_{BS} is the practitioner Black-Scholes delta, T is the time to maturity and V is the VIX index one day before. The neural network structure consists of an input layer with the number of neurons corresponding number of features, 3 hidden layers each with 80 neurons, and an output layer with 1 neuron all fully connected. The activation functions are sigmoid functions after each hidden layer and one linear function at the output layer. The abstract formulation of the neural network is

$$\Phi = W_4 \circ \rho \circ W_3 \circ \rho \circ W_2 \circ \rho \circ W_1 \quad (9)$$

where $W_1 : \mathbb{R}^{3(4)} \rightarrow \mathbb{R}^{80}$, $W_2 : \mathbb{R}^{80} \rightarrow \mathbb{R}^{80}$, $W_3 : \mathbb{R}^{80} \rightarrow \mathbb{R}^{80}$, $W_4 : \mathbb{R}^{80} \rightarrow \mathbb{R}$ are the associated affine transformations

$$W_1(x) := A_1x + b_1 \quad A_1 \in \mathbb{R}^{3(4) \times 80} \quad b_1 \in \mathbb{R}^{80} \quad (10)$$

$$W_2(x) := A_2x + b_2 \quad A_2 \in \mathbb{R}^{80 \times 80} \quad b_2 \in \mathbb{R}^{80} \quad (11)$$

$$W_3(x) := A_3x + b_3 \quad A_3 \in \mathbb{R}^{80 \times 80} \quad b_3 \in \mathbb{R}^{80} \quad (12)$$

$$W_4(x) := A_4x + b_4 \quad A_4 \in \mathbb{R}^{80 \times 1} \quad b_4 \in \mathbb{R} \quad (13)$$

and ρ is the sigmoid function that acts pointwise, namely,

$$\rho(x) = \frac{1}{1 + e^{-x}} \quad (14)$$

By Universal Approximation Theorem, let $K \subset \mathbb{R}^{3(4)}$ be compact

$$\sup_{(r, \delta_{\text{BS}}, T) \in K} \|\Phi(r, \delta_{\text{BS}}, T) - f(r, \delta_{\text{BS}}, T)\| < \epsilon \quad \forall \epsilon > 0 \quad (15)$$

$$\sup_{(r, \delta_{\text{BS}}, T, V) \in K} \|\Phi(r, \delta_{\text{BS}}, T, V) - g(r, \delta_{\text{BS}}, T, V)\| < \epsilon \quad \forall \epsilon > 0 \quad (16)$$

In this study, we use a mini-batch size of 512 and Adam optimizer with Glorot uniform initializer according to Cao et al. [2020]. Decay rates of first and second moments and learning rate are chosen to be 0.9, 0.999, and 10^{-8} , respectively for Adam optimizer as suggested by Kingma and Ba [2014]. Both models are trained for 2000 epochs instead of 4000 epochs as proposed by Cao et al. [2020] because the MinMaxScaler transformation of data has proven to give a faster convergence rate without overfitting and less loss than reported by them.

2.2.3 XGBoost Model

XGBoost is a scalable end-to-end tree boosting machine learning system proposed by Chen and Guestrin [2016]. It has proven efficiency in handling large datasets and was able to produce state-of-the-art results in recognized machine learning challenges. We aim to demonstrate that XGBoost can match or even surpass the performance of the aforementioned neural network architectures in learning implied volatility change. The following sections will detail this comparison and showcase XGBoost's effectiveness.

One enhancement that significantly boosts the performance of this algorithm in tree boosting is the incorporation of a regularized learning objective. For a dataset comprising n samples and m features $\mathcal{D} =$

$\{(\mathbf{x}_i, y_i)\}$ ($|\mathcal{D}| = n, \mathbf{x}_i \in \mathbb{R}^m, y_i \in \mathbb{R}$),

$$\begin{aligned}\mathcal{L}(\phi) &= \sum_i^n l(\hat{y}_i, y_i) + \sum_k^K \Omega(f_k) \quad \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \\ \hat{y}_i &= \phi(\mathbf{x}_i) = \sum_{k=1}^K f_k(\mathbf{x}_i) \quad f_k \in \mathcal{F} = \{f(\mathbf{x}) = w_{q(\mathbf{x})}\}\end{aligned}\tag{17}$$

where

- \mathcal{L} is the objective function subject to ϕ .
- l is a differentiable convex loss function, e.g. means squared loss.
- \hat{y}_i and y_i are estimated and observed target values. In particular \hat{y}_i is estimated by tree boosting as shown in the second line.
- \mathcal{F} is the space of regression trees, with $q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T$, a mapping from dataset to corresponding leaf index. It represents the tree structure.
- Ω is a penalty term that restricts the complexity of the model.
- T is the number of leaves in the tree and w is leaf weight.
- γ and λ are penalty rates.

The objective function is optimized in an additive manner by greedily adding f_t that improves it the most. Using second-order approximation to get the simplified objective function, shown below

$$\begin{aligned}\tilde{\mathcal{L}}^{(t)} &= \sum_{i=1}^n \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t) \\ g_i &= \partial_{\hat{y}_i^{(t-1)}} l(\hat{y}_i^{(t-1)}, y_i) \\ h_i &= \partial_{\hat{y}_i^{(t-1)}}^2 l(\hat{y}_i^{(t-1)}, y_i)\end{aligned}\tag{18}$$

where

- $\tilde{\mathcal{L}}^{(t)}$ is the objective function at step t.
- $\hat{y}_i^{(t-1)}$ is the estimated target value of i -th data point at the t -th iteration.

Chen and Guestrin [2016] addressed the optimization in XGBoost with the Exact Greedy Algorithm and its approximation, efficiently determining optimal decision tree splits. To combat overfitting, they integrated shrinkage, as proposed by Friedman [2002], which moderates

the impact of individual trees, and column subsampling, a technique by Breiman [2001] that enhances model robustness through feature selection. Moreover, XGBoost employs a column block structure for parallel learning, optimizing memory use and computation, especially in large datasets. Its design also accounts for data sparsity, ensuring robust performance even with missing or zero-valued data. In summary, XGBoost’s blend of algorithmic precision and computational efficiency makes it a fitting choice for our study, showcasing its potential against neural network models in our specific task.

The XGBoost model is also trained on the same 3 features as NN-3 and with an additional feature, the VIX index as its NN-4 counterpart. We have selected a learning rate of 0.01 and employed early stopping with a threshold of 20 to effectively mitigate overfitting issues. It performs the step-size shrinkage, and weights relating to the new features can be easily extracted. The number of trees and individual tree depth is restricted to 1000 and 10, respectively for both models, which will be discussed in section 3.

2.2.4 LightGBM Model

In light of the challenges posed by high-dimensional data, LightGBM, introduced by Ke et al. [2017], emerges as an advanced implementation of tree boosting that directly addresses the inefficiencies found in the XGBoost framework. Where XGBoost’s routine to scan all data points for each feature to estimate the information gain of all possible splits becomes time-consuming, LightGBM introduces key innovations that markedly enhance its efficiency and effectiveness, particularly in high-dimensional spaces.

Central to LightGBM’s approach is the Gradient-based One-Side Sampling (GOSS), which smartly prioritizes data points with larger gradients, or errors, thus focusing the learning on the most difficult-to-predict data points. This selective attention not only quickens the training process but also sharpens the accuracy of the model, as it learns more from the most informative data points.

Alongside GOSS, LightGBM employs Exclusive Feature Bundling (EFB), an ingenious solution to the curse of dimensionality. This technique intelligently bundles together features that are rarely non-zero at the same time, effectively reducing the number of features with-

out significantly losing information. Such reduction is crucial in high-dimensional data, as it directly impacts the computational load and efficiency.

Furthermore, LightGBM optimizes the process of finding the best split in trees through a histogram-based approach. This method involves grouping feature values into discrete bins, simplifying the determination of optimal split points. This histogram-based split finding not only reduces the computational expense but also maintains a high level of accuracy in the model.

In our empirical analysis, we will demonstrate that LightGBM has a superior performance over XGBoost, particularly in handling the additional feature VIX index. Models are trained for both feature dimensions - LGMB-3 and LGBM-4. The learning rate and the number of tree restrictions are identical to XGBoost models for comparison purposes. However, we only limit the number of leaves 2^{10} instead of the depths of trees due to LightGBM's leave-wise tree growth nature to control model complexity. We additionally lower bound the minimum number of data points split with 200 to control model complexity to avoid the model being too tailored to the specificities and noise. Early stopping at the 20th round is again required.

2.3 Loss measures

Three loss metrics are used in this study: Mean Squared Error (MSE), Mean Absolute Error (MAE), and coefficient of determination of Mincer-Zarnowitz regression (R^2).

2.3.1 Mean Squared Error

MSE is a measure of the average squared difference between the estimated values and the actual value, defined as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\mathbb{E}[\Delta\sigma_{\text{imp}}]_i - \mathbb{E}[\widehat{\Delta\sigma}_{\text{imp}}]_i)^2 \quad (19)$$

where $\mathbb{E}[\Delta\sigma_{\text{imp}}]_i$ is the actual observed implied volatility change, while $\mathbb{E}[\widehat{\Delta\sigma}_{\text{imp}}]_i$ is the estimated implied volatility change, and n is the size of training, validation or test set. These notations are consistently applied in the subsequent formulas.

Building on the framework from Cao et al. [2020], we define the Gain of model A relative to model B as follows:

$$\text{Gain} = 1 - \frac{\text{model A}}{\text{model B}}$$

2.3.2 Mean Absolute Error

MAE is a measure of the average absolute difference between the estimated values and the actual value, defined as

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\mathbb{E}[\Delta\sigma_{\text{imp}}]_i - \mathbb{E}[\widehat{\Delta\sigma}_{\text{imp}}]_i| \quad (20)$$

2.3.3 Mincer-Zarnowitz R^2

The Mincer-Zarnowitz regression is employed to evaluate the precision of forecast models Pinder [2022]. Ideally, the actual values $\mathbb{E}[\Delta\sigma_{\text{imp}}]$ are equal to the forecasted values $\mathbb{E}[\widehat{\Delta\sigma}_{\text{imp}}]$, i.e.

$$\mathbb{E}[\Delta\sigma_{\text{imp}}] = \mathbb{E}[\widehat{\Delta\sigma}_{\text{imp}}] \Leftrightarrow \mathbb{E}[\Delta\sigma_{\text{imp}}] = 0 + 1 \cdot \mathbb{E}[\widehat{\Delta\sigma}_{\text{imp}}] \quad (21)$$

Therefore, a hypothesis test is formulated as

$$\begin{aligned} \mathbb{E}[\Delta\sigma_{\text{imp}}] &= \beta_0 + \beta_1 \mathbb{E}[\widehat{\Delta\sigma}_{\text{imp}}] \\ H_0 : \beta_0 &= 0 \text{ and } \beta_1 = 1 \\ H_1 : &\text{Otherwise} \end{aligned} \quad (22)$$

The R^2 from the Mincer-Zarnowitz regression serves as valuable metrics for comparing the efficacy of different forecast models, regardless of whether or not the forecasts are constructed using regression.

3. Empirical results

3.1 Data

The empirical study here is based on the S&P 500 Call options data from OptionMetrics between January 2019 and December 2022. For each trading day during this period, the dataset includes the practitioner delta δ_{BS} , time to maturity T , the return of the S&P 500 index r , the VIX Index V , and implied volatility σ_{imp} for each individual option.

After removing the observations with incomplete information, we do the following steps to clean the data:

1. Remove all option data with a time to maturity T of less than 14 days.
2. Remove all data where delta δ_{BS} is less than 0.05 or greater than 0.95.
3. Sort the data so that data for the same option is arranged in successive trading days.
4. Normalize the implied volatility changes for the same option between successive trading days to a daily scale.

Following the data filtering process, the dataset comprises approximately 4.51 million observations for approximately 124 thousand call options. Table 3 is its statistical description:

Table 3 Summary Statistics of S&P 500 Call Options' Practitioner Delta (δ_{BS}) and Time to Maturity (T), S&P 500 Index Return (r), and VIX Index (V)

	δ_{BS}	T	r	vix	$\Delta\sigma_{imp}$
mean	0.61	157	0.02%	23.20	0.04%
std	0.28	212	1.49%	8.69	1.36%
min	0.05	14	-11.98%	11.54	-46.48%
10%	0.17	23	-1.51%	14.80	-0.89%
50%	0.69	93	0.07%	21.89	0.00%
90%	0.92	359	1.47%	31.84	0.98%
max	0.95	1991	9.38%	82.69	50.06%

To ensure a fair comparison across all models in Section 3, the dataset was randomly divided into training, validation, and test sets in a 7 : 2 : 1 ratio. For the HW model, we combined the training and validation sets to form a new training set, whereas for other models, the three sets were kept separate. The models were trained using the training and validation sets and evaluated on the test set.

In Section 5, we initially selected models using the training and validation sets. We then trained and tested the HWR, HWT, HWRT, and HWRTF models on the same dataset used for the HW model in Section 3. To further investigate the models' performance differences

on ATM/OTM and ITM call options, we divided the entire dataset into two subsets based on δ_{BS} values within the ranges $[0.05, 0.5]$ and $(0.5, 0.95]$ respectively. We then repeated the above steps for each subset. Lastly, to explore if the models are also applicable to put options, we got S&P 500 put option data from OptionMetrics for the same period. After data cleaning (unlike call options, for put options we removed those with δ_{BS} greater than -0.05 or less than -0.95), we repeated the data division process. The summary statistics of S&P 500 Put Options are presented in the Table 12.

3.2 Model training

3.2.1 Hull-White Model

The examination of Table 4 reveals that all parameters within the Hull-White (HW) model demonstrate significant statistical weight at a 1% threshold. This significance underscores a robust deviation from the null hypothesis of zero value. In particular, the parameters β_1 and β_3 are greater than zero, while β_2 is less than zero. These findings are consistent with the results obtained by Cao et al. [2020]. Additionally, the integration of an intercept term, denoted as β_0 , along with the distinct measurement methodologies employed for assessing time to maturity *Tour* study adopts a 365-day year as opposed to the 252-day financial year utilized by Cao et al. [2020] introduces quantitative variances in parameter estimations. However, it is noteworthy that, despite these numerical discrepancies, the proportional relationships inherent in the parameter values remain congruent with those reported by them.

Table 4 Estimated Parameters of the Hull-White Model (HW Model) from January 2019 to December 2022: Analysis Using Training and Validation Sets

Parameter	value	std	t-statistic	p-value
β_0	0.000578	0.000006	95.879374	0
β_1	-0.120133	0.001288	-93.262477	0
β_2	0.106329	0.007540	14.102356	0
β_3	-0.241175	0.007951	-30.331639	0

3.2.2 Neural Network Model

In Figure 1, it's observed that following the neural network structure, parameter initialization, and training procedures outlined in Section 2, there's a consistent decrease in Mean Squared Error (MSE) on the training set during the model's training. The MSE on the validation set also decreases, albeit with some fluctuations, indicating a generally downward trend with minor variations, likely due to the implementation of mini-batch gradient descent. These observations suggest a low risk of overfitting, demonstrating the model's robustness.

Moreover, we experimented with dropout and regularization terms as methods to mitigate overfitting. However, it was determined that these measures were not essential for our model, given the substantial size of our dataset and the simplicity of our neural network's topology. Additionally, the implementation of min-max scaling on all data has enhanced the model's convergence speed compared to the results of Cao et al. [2020].

3.2.3 Boosting Model

Figure 1 illustrates that as the number of training iterations increases, all models consistently show a decrease in Mean Squared Error (MSE) on both the training and validation datasets, indicating a minimal risk of model overfitting. While both k-Fold and Bayesian cross-validation methods typically lean towards more complex model configurations, such as a preference for greater depth in the XGBoost model and larger leaf sizes in LightGBM, the parameters listed in Section 2 are justified for the two reasons:

- Although increasing model complexity can lead to marginal improvements in performance, the benefits are often minimal and come with an increased risk of overfitting. This is particularly relevant in our context, where the XGBoost-3 model, under identical model parameters, already exhibits a higher risk of overfitting compared to the XGBoost-4 model, albeit still at a low level. We have included an additional set of results using an XGBoost-

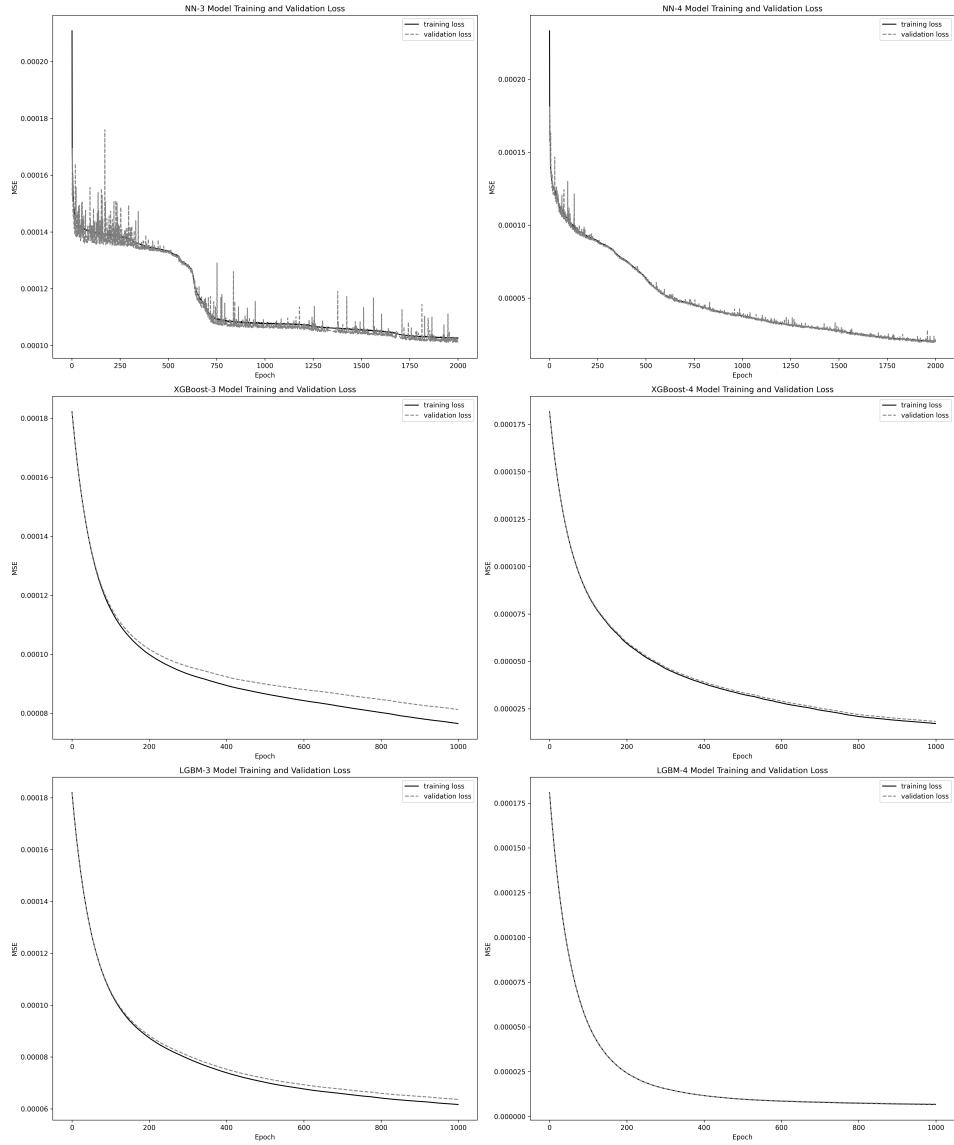


Figure 1. MSE Evolution on Training and Validation Sets Through Epochs for NN-3/4 Model(Up), XGboost-3/4(Middle) and LightGBM-3/-4(Down)

3 model with a maximum depth of 8 in the subsequent model comparison section to provide a comprehensive view.¹

- Under the current hyperparameter settings, boosting models have already demonstrated significantly superior performance compared to neural network models, training more rapidly and showing lower errors on the test dataset.

Consequently, our decision to employ boosting models for examining the features influencing implied volatility movements was based on their satisfactory performance against our criteria.

3.3 Model Comparison

Table 5 Empirical MSE, MAE, R^2 and Gain for HW Model, NN-3/NN-4 Model, XGBoost-3/XGBoost-4 Model and LGBM-3/LGBM-4 Model (in-sample)

	HW	NN-3	XGBoost-3	LGBM-3
MSE	0.000145	0.000102	0.000077	0.000062
MAE	0.006370	0.005630	0.004888	0.004165
R^2	0.208664	0.446352	0.593972	0.669590
Gain		29.66%	46.90%	57.24%
	HW	NN-4	XGBoost-4	LGBM-4
MSE	0.000145	0.000019	0.000017	0.000007
MAE	0.006370	0.002560	0.002239	0.001130
R^2	0.208664	0.894164	0.915619	0.963891
Gain		86.90%	88.28%	95.17%

In accordance with Tables 5 and 6, we make the following observations:

- All machine learning-based models exhibit significant improvements in performance compared to the HW model. Notably, XGBoost and LightGBM models demonstrate more substantial enhancements compared to the neural network model. Specifically,

¹ Additional results for the XGBoost-3 model with a maximum depth of 8 are reported in the model comparison section.

Table 6 Empirical MSE, MAE, R^2 and Gain for Benchmark Model, NN-3/NN-4 Model, XGBoost-3/XGBoost-4 Model and LGBM-3/LGBM-4 Model (out-of-sample)

	HW	NN-3	XGBoost-3	LGBM-3
MSE	0.000146	0.000103	0.000083	0.000065
MAE	0.006374	0.005644	0.004954	0.004224
R^2	0.207278	0.441172	0.557351	0.651961
Gain		29.45%	43.15%	55.48%
	HW	NN-4	XGBoost-4	LGBM-4
MSE	0.000146	0.000020	0.000019	0.000007
MAE	0.006374	0.002566	0.002272	0.001154
R^2	0.207278	0.890977	0.906420	0.960499
Gain		86.30%	86.99%	95.21%

XGBoost-3 and LightGBM-3 achieve improvements of 46.9% and 57.24%, respectively, over the HW model, whereas NN-3 shows a gain of only 29.66%. This highlights the superior performance of XGBoost and LightGBM in this context.

- The four-factor model shows a notable performance improvement over the three-factor model, indicating that changes in market sentiment have a significant impact on implied volatility. In particular, XGBoost-4 and LightGBM-4 achieve gains of 88.28% and 95.17%, respectively, over the HW model, while NN-4 shows an 86.90% improvement. This further underscores the effectiveness of boosting-based models in this context.
- In terms of training speed, the neural network models NN-3 and NN-4 significantly lag behind the other models, taking nearly 30 hours to complete training, while the other models can finish within 10 minutes. This observation highlights the computational disadvantage of neural network models.

In summary, neural network models achieve substantial performance improvements by relaxing the linear structure of the HW model. However, it's worth noting that neural networks, guided by the Universal Approximation Theorem (UAT), possess the capability to approximate all continuous functions. As a result, the predictions generated by neu-

ral networks tend to be exceedingly smooth. While this smoothness can be advantageous in certain contexts, it can also lead to the loss of some distinctive features within the data. In this specific problem characterized by a multitude of training samples and relatively few features, boosting-based methods like XGBoost and LightGBM clearly outperform neural network models. Furthermore, XGBoost and LightGBM models exhibit greater robustness when dealing with data that contains outliers, making them more appealing choices compared to neural network models.

4. Exploring Implied Volatility Dynamics

As mentioned in the introduction, boosting models are known for their superior interpretability. In the subsequent sections, we will gauge the significance of each feature in relation to implied volatility, based on their importance scores within the boosting models. Then, we will analyze the specific manner in which each feature impacts implied volatility dynamics.

4.1 Features Importance

From the Table 7, it is observed that in the LGBM-3/LGBM-4 models, the distribution of feature importance is more uniform compared to the XGBoost-3/XGBoost-4 models. This phenomenon can be attributed to the fact that XGBoost models prioritize the most gain during the tree-building process. As a result, if certain factors exhibit high predictive power, their weights become significantly pronounced. Conversely, LightGBM's leaf-wise growth strategy tends to lead to deeper trees, potentially encompassing a wider range of features, which results in a more evenly distributed feature importance.

Table 7 Parameters Importance for XGBoost-3/XGBoost-4 and LightGBM-3/LightGBM-4 Models

	LGBM-3	LGBM-4	XGBoost-3	XGBoost-4
δ_{BS}	14.98%	21.05%	8.77%	7.24%
T	49.75%	19.55%	18.66%	5.88%
r	35.27%	27.97%	72.57%	37.17%
vix		31.42%		49.72%

Combining insights from all four models, it is evident that the VIX index and the return of the S&P 500 index are the most critical features in studying the movement of implied volatility. Additionally, δ_{BS} and T also play a supportive role in our research.

4.2 Leverage effect

The leverage effect refers to the inverse correlation between changes in the implied volatility of an option and the price movements of its underlying asset. Jay and others have already captured this pattern using principal component analysis. In our study, Table 4 illustrates the relationships among the parameters of the analytical model, which can be summarized as follows:

- β_1 and β_3 are both negative ($\beta_1, \beta_3 < 0$).
- β_2 is positive ($\beta_2 > 0$).
- The absolute value of β_1 is greater than the absolute value of β_2 ($|\beta_1| > |\beta_2|$).

These relationships imply the following inequality:

$$\frac{\beta_1 + \beta_2\delta_{BS} + \beta_3\delta_{BS}^2}{\sqrt{T}} < 0$$

since $0 < \delta_{BS}^2 < \delta_{BS} < 1$. The negative coefficients associated with r signify that a downward movement in the S&P 500 index corresponds to an upward shift in the implied volatility of the S&P 500 call option. This observation aligns with the result of the leverage effect.

In Figure 2, we observe a significant negative correlation between r and $\Delta\sigma_{imp}$ when market volatility is low (vix = 15) and at moderate levels of market volatility (vix = 22). Specifically, when r is less than zero, $\Delta\sigma_{imp}$ tends to be positive, while when r is greater than zero, $\Delta\sigma_{imp}$ tends to be negative.

Furthermore, we also find that the relationship between r and $\Delta\sigma_{imp}$ is influenced by time to maturity (T). Options with longer maturities exhibit flatter r - $\Delta\sigma_{imp}$ lines, while options with shorter maturities show steeper slopes in their r - $\Delta\sigma_{imp}$ lines. This suggests that options with shorter maturities are more sensitive and react more swiftly to price fluctuations in the underlying market compared to options with longer maturities.

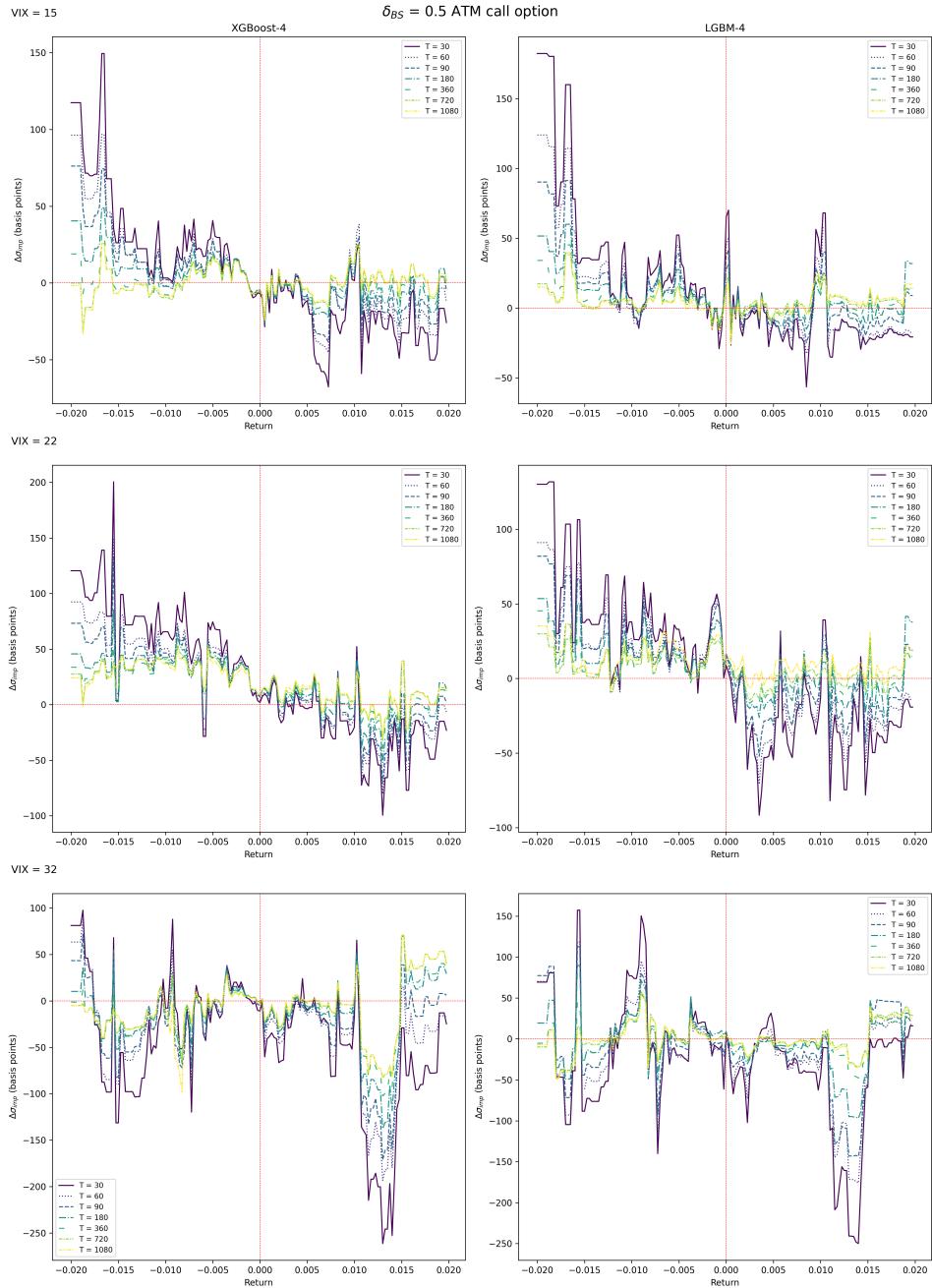


Figure 2. Comparative Analysis of ATM Call Option Implied Volatility ($\Delta\sigma_{\text{imp}}$) Movements and S&P 500 Returns (r) Using XGBoost-4(left) and LightGBM-4 Models(right)

4.3 Sentiment effect

From the Table 5 and 6, After incorporating the Volatility Index (VIX), which reflects market investors' expectations, we observe that all models can better capture the market's movements.

When market fear escalates significantly ($\text{vix}=32$), as observed in Figure 2, Figure 4, and Figure 5, it becomes evident that the degree of the leverage effect diminishes for all moneyness levels of call options. With the increase in VIX, the impact of market sentiment on the $r-\Delta\sigma_{\text{imp}}$ line can be decomposed into the following two components:

- **Rotation:** An increase in VIX causes the $r-\Delta\sigma_{\text{imp}}$ line to rotate, leading to a steeper slope. For options with shorter maturity and larger delta values, this rotation is more significant because these options are more sensitive to changes in market sentiment.
- **Translation:** An increase in VIX shifts the entire $r-\Delta\sigma_{\text{imp}}$ line downwards. This is because in highly panicked market conditions, when security returns are within a moderate range, even if the current S&P 500 returns are positive, market expectations for future returns are negative. Consequently, the demand for call options decreases.

4.4 Inversion effect

In addition to the leverage effect, we observe that when S&P 500 returns rise to a certain level, the slope of the $r-\Delta\sigma_{\text{imp}}$ line undergoes a reversal. This phenomenon occurs because as the underlying asset's price enters a rapid uptrend, the market demand for call options surges, leading to an increase in implied volatility. This effect counteracts the impact of the leverage effect, resulting in a flattening or even a reversal of the $r-\Delta\sigma_{\text{imp}}$ line. Furthermore, during periods of heightened market anxiety ($\text{VIX}=32$), this effect intensifies.

Additionally, by examining the dynamic implied volatility charts generated by XGBoost-3 and LGBM-3 for $r = 1.75\%$ as depicted in Figure 3, or by comparing the variations in the $r-\Delta\sigma_{\text{imp}}$ lines between Figure 4 and Figure 5, we observe that this effect becomes even more pronounced for call options with shorter time to maturity or higher moneyness.

5. Adaptive Analytical Models

5.1 Definitions

Building upon the outcomes of models that integrate four distinct features, we are able to extend the Hull-White model (HW model) by incorporating sentiment effects, encompassing both rotation and translation aspects. Furthermore, this extension process can be advanced by including considerations for inversion effect. The augmented models are formulated as follows:

- HWR(n_1, V_0) Model

$$\mathbb{E}(\Delta\sigma_{\text{imp}}) = \left(\beta_1 \frac{1}{\sqrt{T}} + \beta_2 \frac{\delta_{\text{BS}}}{\sqrt{T}} + \beta_3 \frac{\delta_{\text{BS}}^2}{\sqrt{T}} + \underbrace{\beta_4 \frac{\delta_{\text{BS}} V^{n_1}}{\sqrt{T}} \mathbb{I}_{\{V > V_0\}}}_{\text{Rotation}} \right) r \quad (23)$$

- HWT(n_2, V_0) Model

$$\mathbb{E}(\Delta\sigma_{\text{imp}}) = \left(\beta_1 \frac{1}{\sqrt{T}} + \beta_2 \frac{\delta_{\text{BS}}}{\sqrt{T}} + \beta_3 \frac{\delta_{\text{BS}}^2}{\sqrt{T}} \right) r + \underbrace{\beta_6 V^{n_2} \mathbb{I}_{\{V > V_0\}}}_{\text{Translation}} \quad (24)$$

- HWRT(n_1, n_2, V_0) Model

$$\mathbb{E}(\Delta\sigma_{\text{imp}}) = \left(\beta_1 \frac{1}{\sqrt{T}} + \beta_2 \frac{\delta_{\text{BS}}}{\sqrt{T}} + \beta_3 \frac{\delta_{\text{BS}}^2}{\sqrt{T}} + \underbrace{\beta_4 \frac{\delta_{\text{BS}} V^{n_1}}{\sqrt{T}} \mathbb{I}_{\{V > V_0\}}}_{\text{Rotation}} \right) r + \underbrace{\beta_6 V^{n_2} \mathbb{I}_{\{V > V_0\}}}_{\text{Translation}} \quad (25)$$

- HWRTI(n_1, n_2, V_0, r_0) Model

$$\begin{aligned} \mathbb{E}(\Delta\sigma_{\text{imp}}) &= \left(\beta_1 \frac{1}{\sqrt{T}} + \beta_2 \frac{\delta_{\text{BS}}}{\sqrt{T}} + \beta_3 \frac{\delta_{\text{BS}}^2}{\sqrt{T}} \right) r \\ &+ \left(\underbrace{\beta_4 \frac{\delta_{\text{BS}} V^{n_1}}{\sqrt{T}} \mathbb{I}_{\{V > V_0\}}}_{\text{Rotation}} + \underbrace{\beta_5 \frac{\delta_{\text{BS}}}{\sqrt{T}} \mathbb{I}_{\{r > r_0\}}}_{\text{Inversion}} \right) r + \underbrace{\beta_6 V^{n_2} \mathbb{I}_{\{V > V_0\}}}_{\text{Translation}} \end{aligned} \quad (26)$$

where $n_1 \geq 0$, $n_2 \geq 0$, $r_0 > 0$ and $V_0 > 0$ are constants.

The HWR model accounts for the rotation of the line due to an increase in the VIX, while the HWT model considers the overall translation of the line caused by a rise in the VIX. The HWRT model, on the other hand, addresses both types of changes in the line resulting

from an increase in the VIX. Additionally, the HWRTI model builds upon the HWRT model by also taking into account the line's rotation induced by rapid price increases.

5.2 Model Selection

In equation (23),(24),(25) and (26), we need to determine the values of the parameters n_1 , n_2 , r_0 and V_0 before training the model.

5.2.1 The Range of Parameter

Drawing from the statistical summary of the dataset, we are able to delineate the range for parameter selection as follows:

- (1) Inversion Effect Point: $r_0 \in \{1\%, 1.1\%, 1.2\%, \dots, 2\%\}$;
- (2) Sentiment Effect Point: $V_0 \in \{15, 16, 17, \dots, 32\}$;
- (3) Sentiment Effect Speed: $n_1, n_2 \in \{0, 1, 2\}$;

For r_0 , only about 5% of the returns r exceed 2%. Regarding V_0 , the values 15 and 32 are close to its 10-th and 90-th percentiles, respectively. Consequently, our selection for r_0 and V_0 encompasses nearly all market scenarios.

The selection of the iteration ranges for n_1 and n_2 is primarily based on the following two considerations. On the one hand, constraining the values of n_1 and n_2 to be within the range $\{0, 1, 2\}$ can enhance the stability and generalization capability of the model. On the other hand, the performance improvement of the modified model compared to the benchmark model is the greatest as n_1 and n_2 transition from 0 to 1. As n_1 and n_2 continue to increase, the model's performance continues to improve, but this improvement diminishes rapidly, and in some cases, it can even lead to a deterioration in model performance.

5.2.2 Parameter Selection

Typically, the selection of parameters in such tasks is conducted through the application of k-Fold cross-validation methods. However, considering the specific objectives of this section, which entail the extension of the analytical model and the provision of a reasoned explanation for implied volatility movements, we have chosen to employ an alternative two-step approach:

- (1) We select values for n_1 and n_2 from the set $\{0, 1, 2\}$ while varying V_0 within the range of 15 to 32. Following this, the model is trained using the training dataset, and its performance is assessed against the validation dataset to determine the optimal parameters V_0^* , n_1^* , and n_2^* that minimize the mean squared error (MSE) on the validation dataset.
- (2) Fix V_0^* , n_1^* and n_2^* and vary r_0 within the range of 1.1% to 2.0%, employing the same methodology as in (1) to determine the optimal r_0^* .

From the results presented in Table 14, we have selected the parameters $n_1^* = 2$, $n_2^* = 2$, and $V_0^* = 16$ for the HWR, HWT, and HWRT models. Additionally, for the HWRTI model, we have chosen these parameters along with $r_0^* = 1.5\%$ according to Table 15.

5.3 Model Estimation

From the Table 8, it is evident that all model coefficients are significant at the 1% level, allowing us to infer the characteristics of implied volatility movements for S&P 500 call options as follows:

- (1) **Leverage Effect:** Based on the values of β_1 , β_2 and β_3 extracted from the table for all models, it is apparent that for nearly all S&P 500 call options, the following inequality holds

$$\frac{\beta_1 + \beta_2 \delta_{BS} + \beta_3 \delta_{BS}^2}{\sqrt{T}} < 0$$

since $\delta_{BS} \in [0.05, 0.95]$. This consistently negative slope of the r - $\Delta\sigma_{imp}$ line indicates a significant leverage effect for call options.

- (2) **Sentiment Effect:** Regarding the sentiment effect, our findings show that $\beta_4 < 0$ in all models, leading to

$$\beta_4 \frac{\delta_{BS} V^{n_1}}{\sqrt{T}} < 0$$

since $\delta_{BS} > 0$. This pattern suggests that as the VIX rises, the slope of the r - $\Delta\sigma_{imp}$ line for call options becomes steeper. Moreover, the fact that $\beta_6 < 0$ in the HWT, HWRT, and HWRTI models indicates that an increase in the VIX corresponds to a downward shift in the r - $\Delta\sigma_{imp}$ relationship.

(3) **Inversion Effect:** In the HWRTI model, $\beta_5 > 0$ implies

$$\beta_5 \frac{\delta_{\text{BS}}}{\sqrt{T}} \mathbb{I}_{\{r > r_0\}} > 0.$$

This indicates that when the S&P 500 index begins to rise rapidly, the effect of an increasing VIX on the slope of the r - $\Delta\sigma_{\text{imp}}$ line will be partially offset.

Table 8 Parameter Estimation, Robust Standard Errors (Heteroskedasticity Consistent Covariance Matrix Estimator), and Empirical MSE, MAE, R^2 , and Gain for HW, HWR, HWT, HWRT, and HWRTI Models based on S&P 500 Call Options

	HW	HWR	HWT	HWRT	HWRTI
β_0	0.000578 (0.000006)	0.000505 (0.000006)	0.001618 (0.000019)	0.001220 (0.000014)	0.001443 (0.000013)
β_1	-0.120133 (0.001288)	-0.134666 (0.001188)	-0.123259 (0.001246)	-0.136395 (0.001153)	-0.139615 (0.001114)
β_2	0.106329 (0.007540)	0.451899 (0.007467)	0.105141 (0.007417)	0.441183 (0.007386)	0.423042 (0.007198)
β_3	-0.241175 (0.007951)	-0.488359 (0.007608)	-0.235756 (0.007836)	-0.477558 (0.007563)	-0.512427 (0.007399)
β_4		-0.756180 (0.006980)		-0.734514 (0.006808)	-0.830526 (0.006816)
β_5					0.168416 (0.001720)
β_6			-0.017877 (0.000362)	-0.012266 (0.000273)	-0.024679 (0.000273)
In-sample					
MSE	0.000145	0.000135	0.000144	0.000135	0.000132
MAE	0.006370	0.006143	0.006407	0.006167	0.006115
R2	0.208664	0.263799	0.215789	0.267108	0.281872
Gain		6.97%	0.90%	7.39%	9.25%
Out-of-Sample					
MSE	0.000146	0.000136	0.000145	0.000135	0.000132
MAE	0.006374	0.006149	0.006412	0.006172	0.006121
R2	0.207278	0.263477	0.214209	0.266725	0.280732
Gain		7.09%	0.87%	7.50%	9.26%

Moreover, we know that all adaptive analytical models exhibit certain degrees of improvement compared to the HW model. Specifically, the HWR model shows a 7.09% improvement over the HW model, while the HWT model registers a modest 0.87% enhancement. This suggests that the impact of the r-sigma line's rotation, caused by an increase in VIX, is more significant than its horizontal shift. The HWRT model, which considers both these effects, therefore achieves a higher Gain of 7.50%. Finally, the HWRTI model, incorporating both sentiment effect and inversion effect, attains the highest Gain of 9.26%. In summary, when comparing the models, we see that the rotation of the r-sigma line, caused by a rise in VIX, has a greater effect than inversion effect. The horizontal shift of this line, also due to an increase in VIX, has the least impact.

5.4 Robustness Checks

In this section, we conduct a robustness assessment of our models from two key perspectives. Firstly, we compare the performance of HW, HWRT and HWRTI models using S&P 500 ATM/OTM Call options ($\delta_{BS} \in [0.05, 0.5]$) with those utilizing S&P 500 ITM Call options ($\delta_{BS} \in (0.5, 0.95]$). Secondly, we will train and test HW, HWRT and HWRTI models using the dataset of S&P 500 Put options and compare their parameter estimations and model performance to those established on the call option dataset.

5.4.1 Moneyness

The analysis of the Table 9 reveals a consistent pattern in the coefficients' signs across all models applied to S&P 500 call options, regardless of their moneyness. This consistency underscores the stable presence of the leverage effect, sentiment effect, and inversion effect across different moneyness levels.

Furthermore, a distinct performance advantage is observed for all models when applied to S&P 500 Out-of-the-Money (OTM) and At-the-Money (ATM) call options, as opposed to In-the-Money (ITM) call options. This performance gap is primarily attributed to the inferior liquidity of ITM call options and their relatively muted response to market volatility changes.

Additionally, the enhancements in performance with the HWRT and HWRTI models are more pronounced for OTM/ATM call options than across the spectrum of all call options. Notably, the Gain for the HWRT model has increased from 7.50% to 10.56%, while the HWRTI model shows an increase in Gain from 9.26% to 13.16%. However, it is important to note a slight reduction in Gains for both the HWRT and HWRTI models when applied to ITM call options.

Table 9 Parameter Estimation, Robust Standard Errors (Heteroskedasticity Consistent Covariance Matrix Estimator), and Empirical MSE, MAE, R^2 , and Gain for HW, HWR, HWT, HWRT, and HWRTI Models based on S&P 500 ATM/OTM and ITM Call Options

Model	HW		HWRT		HWRTI	
	A/O	I	A/O	I	A/O	I
β_0	-0.00082087 (0.00000733)	0.00124717 (0.00000817)	0.00064108 (0.00001716)	0.00127346 (0.00002056)	0.00091494 (0.00001667)	0.00156948 (0.00002062)
β_1	-0.15687223 (0.00196189)	-0.29752509 (0.02726111)	-0.15937500 (0.00176070)	-0.32984319 (0.02240202)	-0.16231059 (0.00168495)	-0.28863422 (0.02154148)
β_2	0.47182340 (0.01965452)	0.57321042 (0.07678657)	0.65431789 (0.01751159)	0.98439134 (0.06371034)	0.58473341 (0.01655587)	0.84325139 (0.06140238)
β_3	-0.91144257 (0.03844583)	-0.54205929 (0.05268358)	-0.90966461 (0.03410744)	-0.84205754 (0.04402740)	-0.92110030 (0.03173375)	-0.78339593 (0.04252923)
β_4			-0.69880914 (0.00741058)	-0.75581123 (0.00812506)	-0.76707185 (0.00776049)	-0.83654973 (0.00810103)
β_5					0.27741978 (0.00479665)	0.13928855 (0.00186702)
β_6			-0.02208864 (0.00032619)	-0.00234918 (0.00042428)	-0.03364314 (0.00033092)	-0.01675500 (0.00043945)
In-sample						
MSE	0.00006957	0.00018014	0.00006262	0.00016730	0.00006053	0.00016497
MAE	0.00419118	0.00740735	0.00399911	0.00717901	0.00401966	0.00711099
R2	0.28616811	0.19601860	0.35743261	0.25334567	0.37890109	0.26370858
Gain			9.99%	7.13%	12.99%	8.42%
Out-of-sample						
MSE	0.00007077	0.00017984	0.00006330	0.00016745	0.00006114	0.00016495
MAE	0.00420754	0.00741412	0.00399916	0.00718392	0.00401738	0.00711399
R2	0.28968491	0.19776633	0.36467054	0.25302275	0.38629581	0.26414919
Gain			10.56%	6.89%	13.61%	8.28%

5.4.2 Option Type

From the Table 10, we can deduce the characteristics of the implied volatility movement for S&P 500 put options as follows:

- (1) **Leverage Effect is still significant for put option.** we observe that for non-deep In-the-Money (ITM) put options, the following equation still holds true:

$$\frac{\beta_1 + \beta_2 \delta_{BS} + \beta_3 \delta_{BS}^2}{\sqrt{T}} < 0$$

since $\delta_{BS} \in [-0.95, -0.05]$. This indicates that the leverage effect remains significant for put options, albeit at a lower magnitude compared to call options.

- (2) **The Sentiment Effect has an opposite impact on put options compared to its effect on call options.** Regarding the sentiment effect, our findings indicate that $\beta_4 < 0$ across all models. Consequently, this leads to:

$$\beta_4 \frac{\delta_{BS} V^{n_1}}{\sqrt{T}} > 0$$

since $\delta_{BS} < 0$. This pattern implies that with a rising VIX, the slope of the $r-\Delta\sigma_{imp}$ line for put options tends to become flatter, which contrasts with the behavior observed in call options. Furthermore, the observation that $\beta_6 > 0$ in both the HWT and HWRT models suggests that an increase in the VIX is associated with an upward shift in the $r-\Delta\sigma_{imp}$ relationship. However, in the HWRTI model, β_6 is negative. This is attributed to the model's consideration of inversion effect, which causes the line in the region where $r > 0$ to rotate upwards. Additionally, the impact of inversion effect on the model is greater than the overall movement of the line caused by an increase in the VIX.

- (3) **The impact of the Inversion Effect on put options is opposite to its impact on call options.** In the HWRTI model, $\beta_5 < 0$ which implies

$$\beta_5 \frac{\delta_{BS}}{\sqrt{T}} \mathbb{I}_{\{r > r_0\}} > 0.$$

This indicates that when the S&P 500 index begins to rise rapidly, the flattening of the $r-\Delta\sigma_{\text{imp}}$ line for put options is further accentuated.

Additionally, from the table in the appendix, we find that similar to call options, all models perform significantly better on ATM and OTM S&P 500 put options compared to ITM S&P 500 put options.

Table 10 Parameter Estimation, Robust Standard Errors (Heteroskedasticity Consistent Covariance Matrix Estimator), and Empirical MSE, MAE, R^2 , and Gain for HW, HWR, HWT, HWRT, and HWRTI Models based on S&P 500 Put Options

	HW	HWR	HWT	HWRT	HWRTI
β_0	0.00055687 (0.00000690)	0.00057368 (0.00000673)	-0.00116369 (0.00002350)	-0.00086813 (0.00001960)	-0.00058245 (0.00001683)
β_1	-0.09985491 (0.00138174)	-0.08539352 (0.00145507)	-0.09888936 (0.00133699)	-0.08533311 (0.00140845)	-0.08677104 (0.00131587)
β_2	-0.13418444 (0.00830927)	0.06776819 (0.00966070)	-0.13197667 (0.00793680)	0.05916825 (0.00917736)	0.2468304 (0.00868362)
β_3	-0.02376189 (0.00968926)	-0.02092625 (0.00962414)	-0.02040535 (0.00927123)	-0.01826192 (0.00918345)	-0.01806771 (0.00813622)
β_4		-0.60329125 (0.00936007)		-0.57207727 (0.00868244)	-0.55331163 (0.00696123)
β_5					-0.41643063 (0.00315455)
β_6			0.02728223 (0.00044462)	0.02284843 (0.00037992)	-0.01338033 (0.00035529)
In-sample					
MSE	0.00018285	0.00017250	0.00017912	0.00016991	0.00015560
MAE	0.00583628	0.00595707	0.00593387	0.00601194	0.00601570
R2	0.02623806	0.08135585	0.04610675	0.09514380	0.17135924
Gain		5.66%	2.04%	7.08%	14.90%
Out-of-Sample					
MSE	0.00018309	0.00017264	0.00017914	0.00016979	0.00015512
MAE	0.00584154	0.00595882	0.00593960	0.00601264	0.00601147
R2	0.02718943	0.08275388	0.04820298	0.09785576	0.17586637
Gain		5.71%	2.16%	7.26%	15.28%

6. Conclusion

In our study, machine learning, traditionally used for prediction and classification, was applied to investigate the relationships among op-

tion implied volatility movements, S&P 500 index returns, the time to maturity of options, practitioner delta, and market sentiment. Our findings confirm the often negative correlation between underlying asset returns and implied volatility changes, as posited by Cont and Da Fonseca [2002], and further uncover the nuances of sentiment and inversion effects. These effects modulate this negative correlation, with variations contingent upon the Volatility Index and the rate of asset returns.

Our approach involved constructing models using XGBoost and LightGBM, two methods based on boosting techniques, and comparing them with analytical model from Hull and White [2017] and neural network model from Cao et al. [2020]. Given the large sample size and limited feature set of our dataset, the XGBoost and LightGBM models excelled over the neural network model in training efficiency, performance, and robustness against outliers. Importantly, these models facilitated the quantification of feature importance scores, essential for deeper analysis into how various features influence implied volatility movements.

We discovered three pivotal effects impacting implied volatility movements: leverage, sentiment, and inversion effects. To explore these, we developed adaptive analytical models based on Hull's original framework. These models included the HWR, which considers the rotation of the $r-\Delta\sigma_{\text{imp}}$ line as VIX increases, and the HWT, focusing on the translation of the line. The HWRT model integrates both rotation and translation effects. Additionally, the HWRTI model, while incorporating the sentiment effect, also considers the diminishing impact of the leverage effect due to high underlying asset returns, a phenomenon we term the inversion effect. The robustness of our models was verified against various scenarios involving option moneyness and type, demonstrating enhanced interpretability and performance compared to the analytical model put produced by Hull and White [2017].

In summary, the characteristics of implied volatility movements are as follows:

- **Leverage Effect:** A predominantly negative correlation between asset returns and implied volatility changes for both call and put options, with the degree of correlation diminishing as time to maturity increases. This indicates longer-dated options are less sensitive to short-term changes in underlying asset returns.

- **Sentiment Effect:** As the VIX rises, indicating heightened market risk expectations and moderate asset returns, the $r-\Delta\sigma_{imp}$ line slope for call options steepens and shifts downwards, and vice versa for put options. This is attributed to the increased demand for put options and decreased demand for call options even with positive asset returns during high market anxiety. Additionally, this effect is more pronounced in shorter-dated options.
- **Inversion Effect:** Higher asset returns, raising expectations of rapid future price increases, lead to increased demand for call options and a flattening or reversal of the $r-\Delta\sigma_{imp}$ line slope. For put options, the slope also flattens, which interferes with the sentiment effect's line shift. It becomes unclear which effect predominates when both returns and market anxiety are high.

In our forthcoming work, we aim to enhance the adaptive analytical models through two strategic approaches. Firstly, we plan to integrate the k-fold cross-validation technique into our model selection process. This method will enable us to fine-tune the model parameters, thereby optimizing the overall performance of the models. Secondly, we intend to expand our research scope by incorporating a broader range of features that influence the movements of implied volatility. By considering these additional variables, we anticipate further refinement and improvement of our models, leading to more comprehensive and accurate analyses of implied volatility dynamics

7. Extension

Extending our models can be approached by incorporating additional factors, such as macro data. In a study by Tuysuzoglu [2008], they highlight that unexpected announcement figures for the core Consumer Price Index (CPI) also influence the behavior of the VIX significantly. CPI gauges inflation in the price of consumer goods and is expected to have a significant impact on asset values. As prices of consumer goods increase, the purchasing power of individuals decreases, causing the economy to contract and equity prices to decline. Monthly changes in CPI are obtained from the Bureau of Labor Statistics as seasonally adjusted rates.

Since the VIX is a crucial quantity when modeling the change in implied volatility, it could make sense to include the CPI as a fifth feature. Using the same machine learning pipeline as in Section 2.2.2, we implemented two 5-feature models by adding the monthly CPI and the monthly relative change in CPI to the daily data and therefore $x := (r, \delta_{BS}, T, V, CPI)$. Due to obvious computational complexity issues, we only included data between 2018-2022 and used 1000 epochs. We obtained the following gains with respect to the HW model:

Table 11 Performance Comparison of 5-Feature and 4-Feature Models Over 1000 Epochs

Model	Gain
4-Feature Model	79.77%
5-Feature Model (CPI)	34.68%
5-Feature Model ($\frac{\Delta CPI}{CPI}$)	91.73%

In Table 11, one can observe that $\frac{\Delta CPI}{CPI}$ as fifth factor is indeed able to improve the 4-feature model. Therefore, jumps in CPI provide crucial information for modeling the change in implied volatility. Still, one has to mention two crucial issues in our method. Firstly, this method may provide some look-ahead bias as the CPI numbers are not announced right at the start of the month and, secondly, since we use data between February 2018 and December 2022 in our 5-feature model, we only use an additional amount of 59 data points which could explain why the CPI number alone reduces the gain of our model.

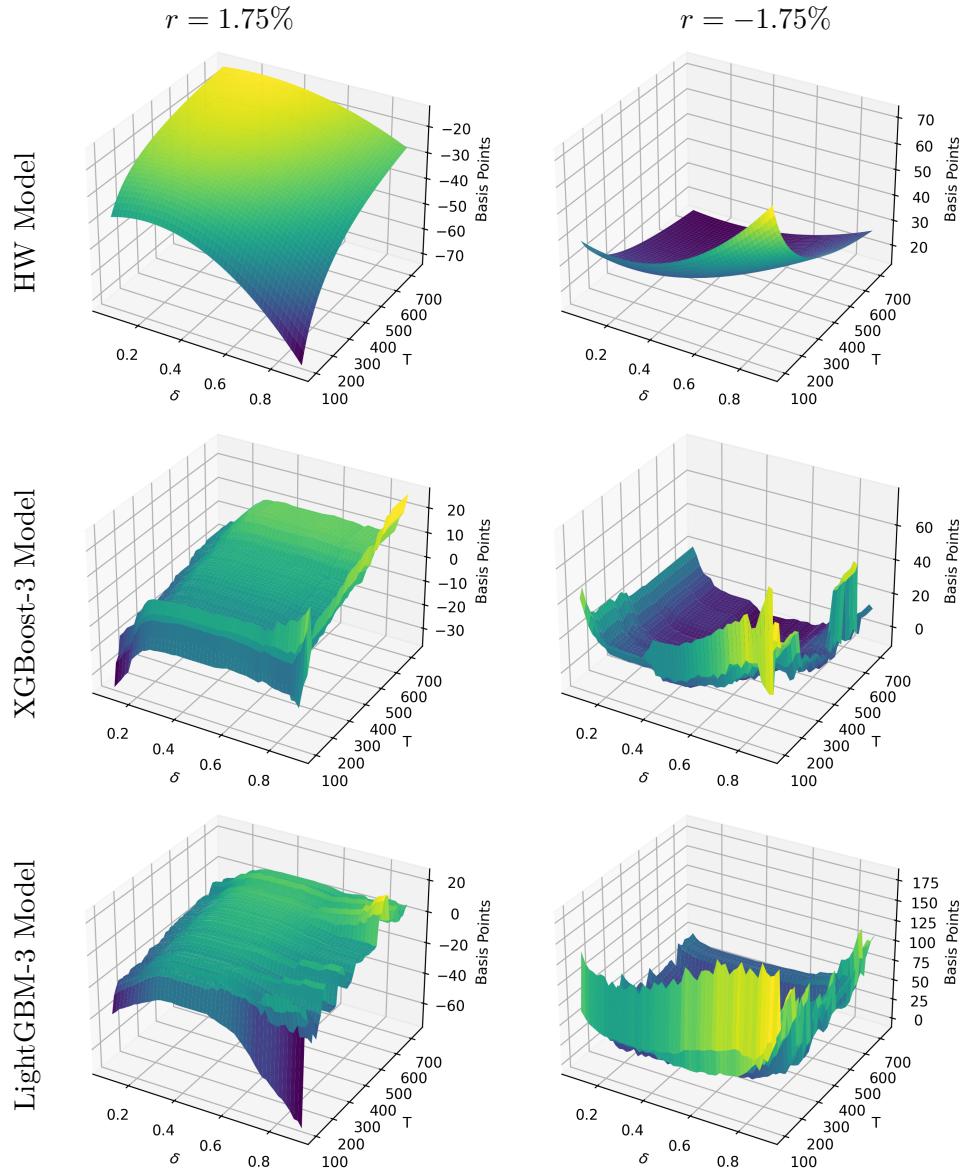


Figure 3. Expected change of implied volatility surface for HW Model, XGBoost-3 Model and LightGBM-3 Model

Appendix

Table 12 Summary Statistics of S&P 500 Put Options' Practitioner Delta (δ_{BS}) and Time to Maturity (T), S&P 500 Index Return (r), and VIX Index (V)

	δ_{BS}	T	r	vix	$\Delta\sigma_{imp}$
mean	-0.39	152	0.04%	23.69	0.05%
std	0.28	205	1.59%	9.38	1.37%
min	-0.95	14	-11.98%	11.54	-27.77%
10%	-0.85	22	-1.55%	14.91	-0.77%
50%	-0.32	91	0.08%	22.10	0.01%
90%	-0.08	349	1.50%	32.45	0.75%
max	-0.05	1991	9.38%	82.69	70.58%

Table 13 Model Selection: Determine the n_1^* , n_2^* and V_0^* for the HWR, HWT, HWRT and HWRTI models utilizing the S&P 500 Call option dataset

(n_1, n_2)	(0,1)	(0,2)	(1,0)	(1,1)
15	0.00014406	0.00014338	0.00013727	0.00013697
16	0.00014387	0.00014317	0.00013717	0.00013690
17	0.00014388	0.00014316	0.00013735	0.00013714
18	0.00014381	0.00014315	0.00013754	0.00013732
19	0.00014417	0.00014343	0.00013802	0.00013785
20	0.00014414	0.00014342	0.00013838	0.00013818
21	0.00014427	0.00014354	0.00013867	0.00013853
22	0.00014421	0.00014350	0.00013903	0.00013885
23	0.00014393	0.00014327	0.00013909	0.00013889
24	0.00014385	0.00014321	0.00013921	0.00013901
25	0.00014375	0.00014310	0.00013914	0.00013897
26	0.00014340	0.00014279	0.00013879	0.00013873
27	0.00014306	0.00014245	0.00013876	0.00013866
28	0.00014362	0.00014299	0.00013974	0.00013955
29	0.00014302	0.00014241	0.00013925	0.00013904
30	0.00014255	0.00014199	0.00013904	0.00013878
31	0.00014244	0.00014192	0.00013918	0.00013891
32	0.00014249	0.00014194	0.00013933	0.00013907
(n_1, n_2)	(1,2)	(2,0)	(2,1)	(2,2)
15	0.00013652	0.00013502	0.00013474	0.00013439
16	0.00013644	0.00013495	0.00013470	0.00013436
17	0.00013666	0.00013499	0.00013477	0.00013440
18	0.00013688	0.00013491	0.00013472	0.00013439
19	0.00013735	0.00013513	0.00013497	0.00013458
20	0.00013769	0.00013518	0.00013501	0.00013463
21	0.00013802	0.00013530	0.00013517	0.00013478
22	0.00013835	0.00013539	0.00013525	0.00013487
23	0.00013840	0.00013540	0.00013525	0.00013488
24	0.00013852	0.00013543	0.00013528	0.00013492
25	0.00013849	0.00013546	0.00013534	0.00013498
26	0.00013829	0.00013532	0.00013532	0.00013500
27	0.00013820	0.00013536	0.00013531	0.00013498
28	0.00013905	0.00013590	0.00013578	0.00013541
29	0.00013855	0.00013569	0.00013555	0.00013518
30	0.00013832	0.00013561	0.00013541	0.00013505
31	0.00013848	0.00013576	0.00013556	0.00013521
32	0.00013861	0.00013593	0.00013573	0.00013538

Table 14 Model Selection: Determine the r_0^* for HWRTI models ($n_1 = 2, n_2 = 2, V_0 = 16$) utilizing the S&P 500 Call option dataset

r_0	1.10%	1.20%	1.30%	1.40%	1.50%
MSE(Validation)	0.00013171	0.00013218	0.00013211	0.00013184	0.00013154
r_0	1.70%	1.80%	1.90%	2.00%	
MSE(Validation)	0.00013214	0.00013238	0.00013231	0.00013259	

Table 15 Model Selection: n_1^* , n_2^* , V_0^* and r_0^* for All Models Using S&P 500 Call/Put Option Dataset

Option Type	Call Option			Put Option		
	All	A/O	I	All	A/O	I
n1	2	2	2	2	2	2
n2	2	2	2	2	2	2
vix0	16	15	16	15	30	15
r0	1.50%	1.50%	1.50%	0.50%	0.50%	0.50%

Table 16 Parameter Estimation, Robust Standard Errors (Heteroskedasticity Consistent Covariance Matrix Estimator), and Empirical MSE, MAE, R^2 , and Gain for HW, HWR, HWT, HWRT, and HWRTI Models based on S&P 500 ATM/OTM and ITM Put Options

Model	HW		HWRT		HWRTI	
	A/O	I	A/O	I	A/O	I
β_0	0.000342 (0.00000600)	0.000949 (0.00001700)	0.00015834 (0.00000475)	-0.00185258 (0.00003661)	-0.00051746 (0.00000885)	-0.00221035 (0.00003311)
β_1	-0.058158 (0.00207600)	-1.754369 (0.03509900)	-0.05846348 (0.00204155)	-1.12988743 (0.02816452)	-0.05391163 (0.00194614)	-1.07441574 (0.02417114)
β_2	0.172331 (0.01982700)	-4.849563 (0.09831100)	0.15122793 (0.01936030)	-2.91642777 (0.07964637)	0.43044616 (0.01811339)	-2.56219613 (0.06887711)
β_3	0.346234 (0.03827600)	-3.253421 (0.06614500)	0.39565736 (0.03763726)	-2.07202247 (0.05484329)	0.55112384 (0.03410734)	-1.98322014 (0.04763522)
β_4			0.1677426 (0.00682380)	-0.62004869 (0.00881191)	0.17652114 (0.00684554)	-0.61129534 (0.00732243)
β_5					-0.43571032 (0.00487026)	-0.45790328 (0.00389558)
β_6			0.0082239 (0.00029699)	0.03825909 (0.00063873)	-0.00663764 (0.00031298)	-0.01784568 (0.00064884)
In-sample						
MSE	0.00007600	0.00038600	0.00007529	0.00034273	0.00007211	0.00030827
MAE	0.00440600	0.00881900	0.00438670	0.00888701	0.00441331	0.00894072
R2	0.09147000	0.01682900	0.09655553	0.12671734	0.13480870	0.21453252
Gain			0.93%	11.21%	5.12%	20.14%
Out-of-sample						
MSE	0.00007700	0.00040700	0.00007696	0.00035462	0.00007349	0.00031682
MAE	0.00444800	0.00885300	0.00442949	0.00889984	0.00445592	0.00895596
R2	0.09239700	0.01784700	0.09623613	0.14453476	0.13714887	0.23582191
Gain			0.05%	12.87%	4.56%	22.16%

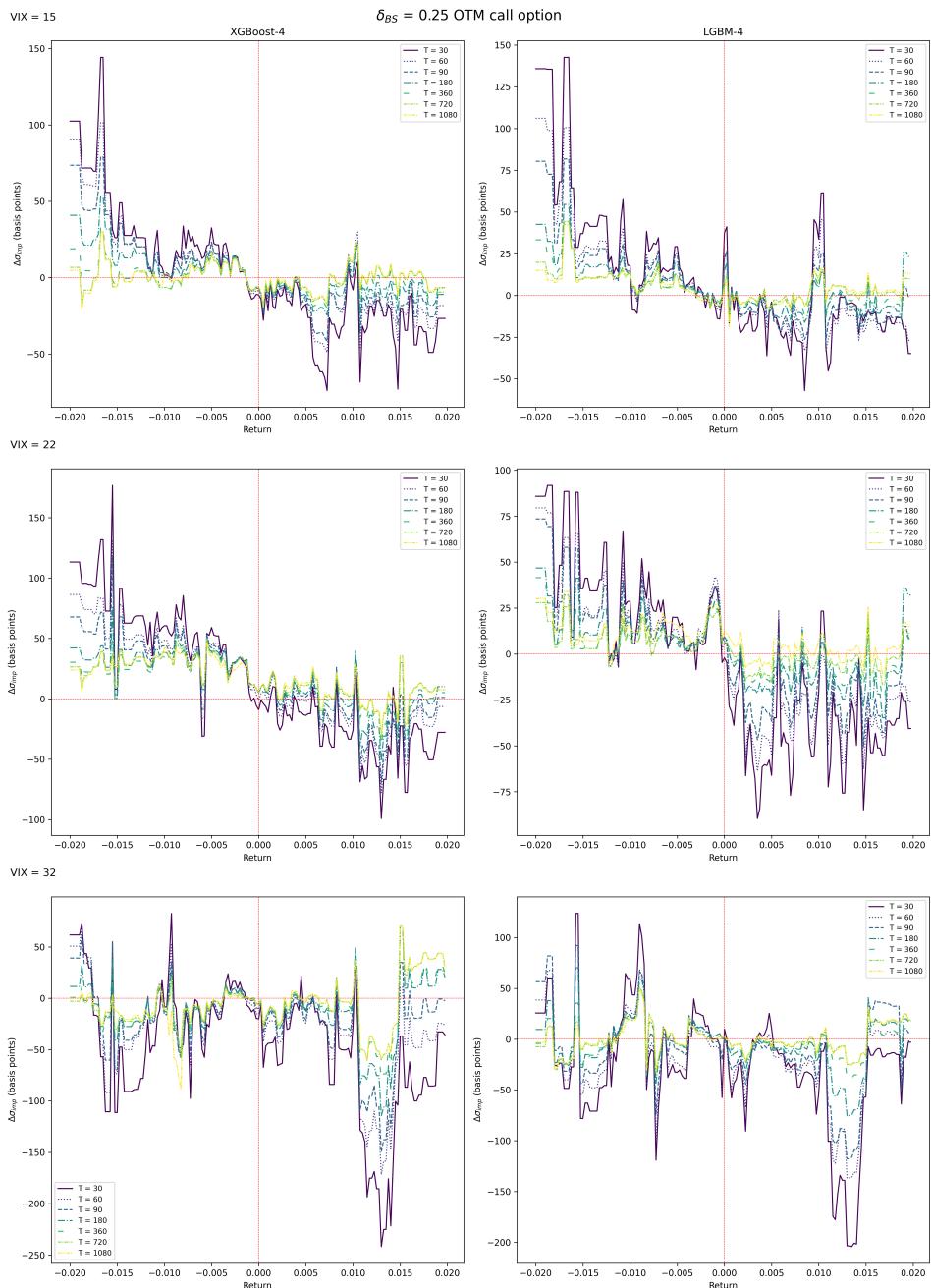


Figure 4. Comparative Analysis of OTM Call Option Implied Volatility ($\Delta\sigma_{\text{imp}}$) Movements and S&P 500 Returns (r) Using XGBoost-4(left) and LightGBM-4 Models(right)

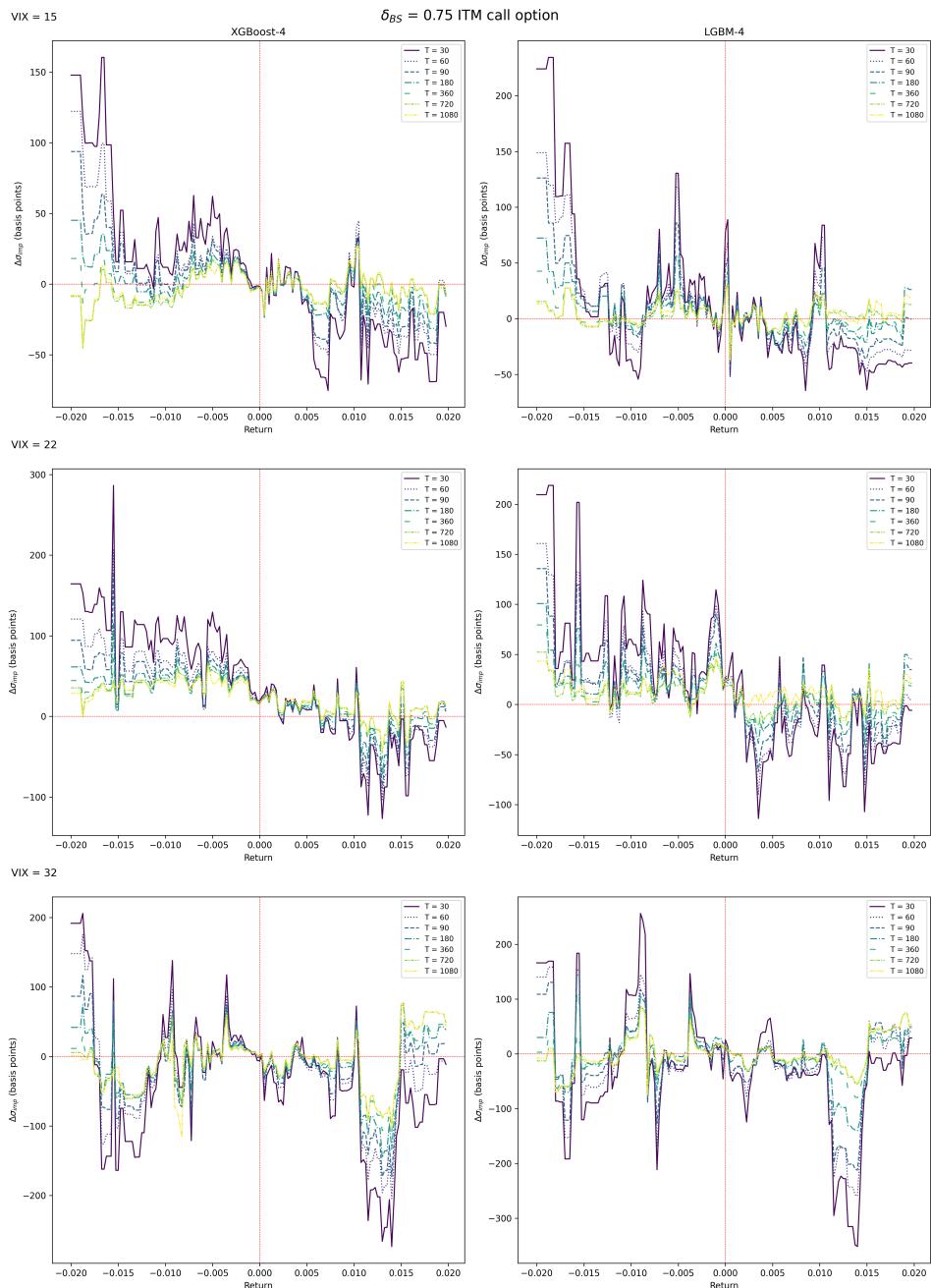


Figure 5. Comparative Analysis of ITM Call Option Implied Volatility ($\Delta\sigma_{\text{imp}}$) Movements and S&P 500 Returns (r) Using XGBoost-4(left) and LightGBM-4 Models(right)

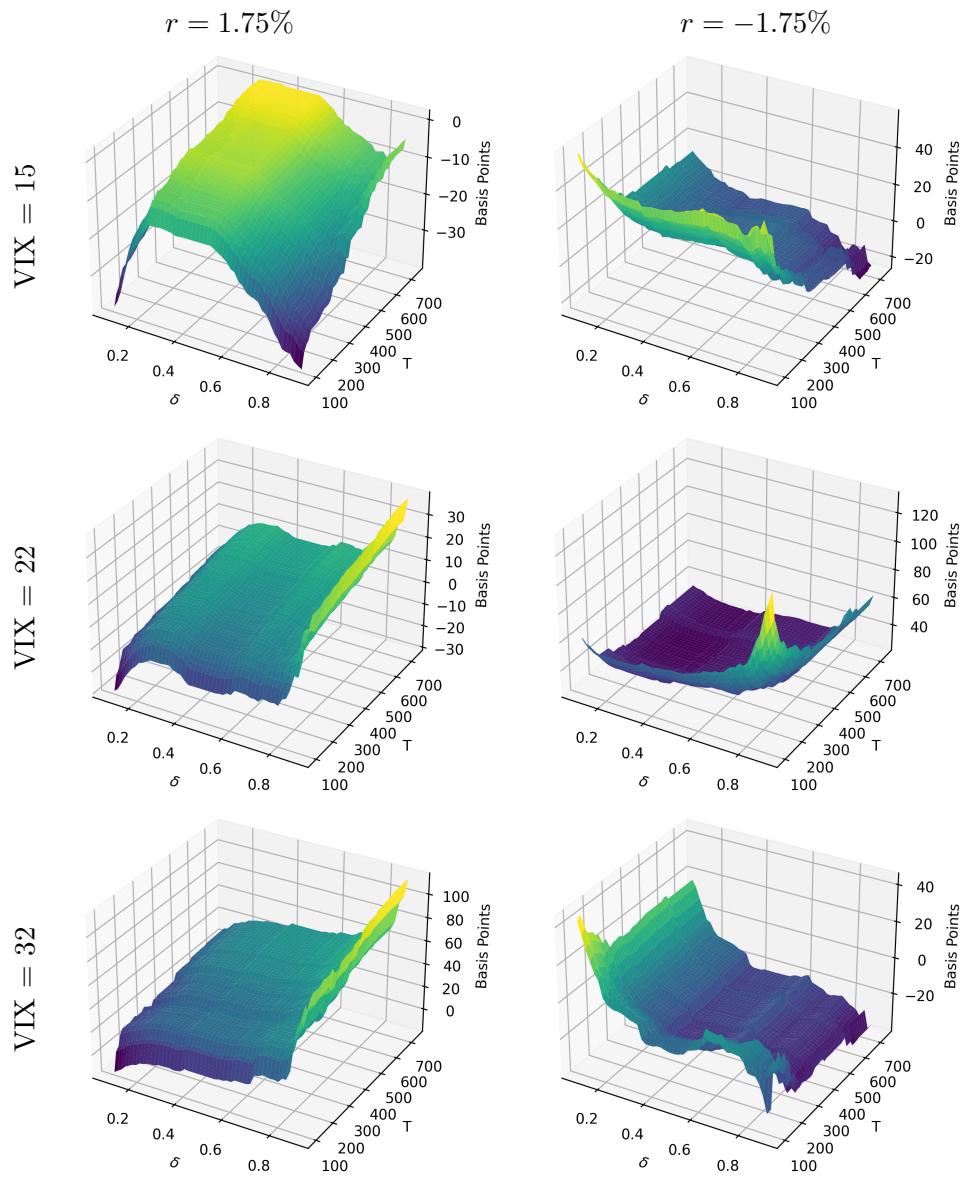


Figure 6. Expected change of implied volatility surface for XGBoost-4 Model

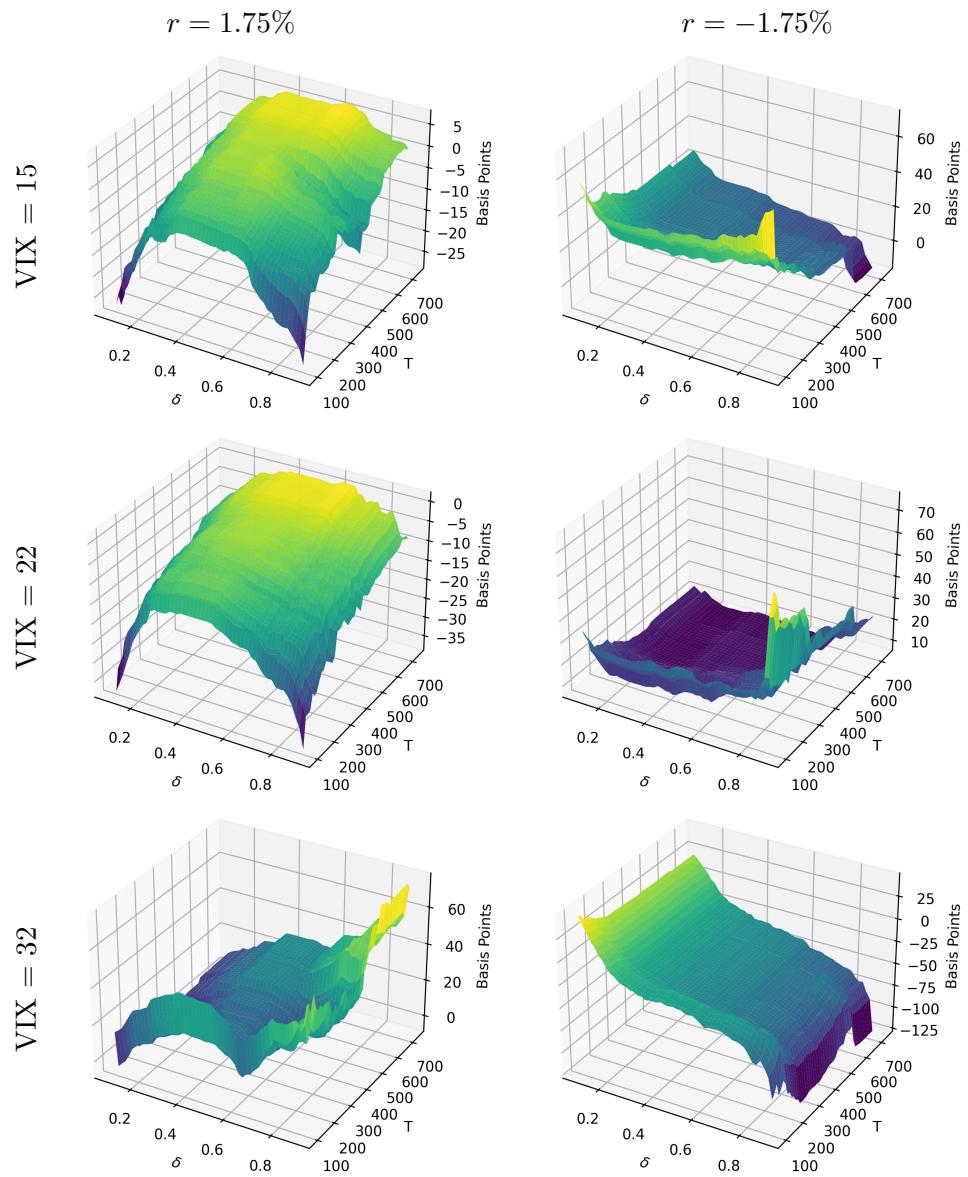


Figure 7. Expected change of implied volatility surface for LightGBM-4 Model

References

- 1 Geert Bekaert and Guojun Wu. Asymmetric volatility and risk in equity markets. *The review of financial studies*, 13(1):1–42, 2000.
- 2 Fischer Black. Studies of stock price volatility changes. In *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, pages 177–181, Washington DC, 1976. American Statistical Association.
- 3 Tim Bollerslev, Julia Litvinova, and George Tauchen. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 4(3):353–384, 2006.
- 4 Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.
- 5 John Y Campbell and Ludger Hentschel. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of financial Economics*, 31(3):281–318, 1992.
- 6 Jay Cao, Jacky Chen, and John Hull. A neural network approach to understanding implied volatility movements. *Quantitative Finance*, 20(9):1405–1413, 2020.
- 7 Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- 8 Andrew A Christie. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, 10(4):407–432, 1982.
- 9 Rama Cont and José Da Fonseca. Dynamics of implied volatility surfaces. *Quantitative finance*, 2(1):45, 2002.
- 10 Kenneth R French, G William Schwert, and Robert F Stambaugh. Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29, 1987.
- 11 Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.
- 12 Jasmina Hasanhodzic and Andrew W Lo. On blacks leverage effect in firms with no leverage. *The Journal of Portfolio Management*, 46(1):106–122, 2019.
- 13 Thorsten Hens and Sven C Steude. The leverage effect without leverage. *Finance Research Letters*, 6(2):83–94, 2009.
- 14 John Hull and Alan White. Optimal delta hedging for options. *Journal of Banking & Finance*, 82:180–190, 2017.

- 15 Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.
- 16 Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- 17 Jonathan P Pinder. *Introduction to business analytics using simulation*. Academic Press, 2022.
- 18 Rolf Poulsen, Klaus Reiner Schenk-Hoppé, and Christian-Oliver Ewald. Risk minimization in stochastic volatility models: model risk and empirical performance. *Quantitative Finance*, 9(6):693–704, 2009.
- 19 Levent Tuysuzoglu. *The Effects of Macroeconomic Data Surprises on Implied Volatility*. PhD thesis, Stern School of Business New York, 2008.