Modeling Implied Volatility Using Out-of-The-Money Options: A Machine Learning Approach for AAPL

MSc Research Project
MSc in FinTech Information

Laura Gomez
Student ID: x23369795

School of Computing
National College of Ireland

Supervisor: Dr. Brian Byrne

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Laura Gomez
Student ID:	x23369795
Programme:	MSc in FinTech Information
Year:	2025
Module:	MSc Research Project
Supervisor:	Dr. Brian Byrne
Submission Due Date:	15/09/2025
Project Title:	Modeling Implied Volatility Using Out-of-The-Money Options: A Machine Learning Approach for AAPL
Word Count:	6020
Page Count:	18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	*Laura Gomez*
Date:	11th September 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Modeling Implied Volatility Using Out-of-The-Money Options: A Machine Learning Approach for AAPL

Laura Gomez
x23369795

Link to Implementation in Colab

Abstract

This research investigates whether OTM call and put options alone can accurately model Apple Inc. (AAPL) implied volatility surface, and compares predictive accuracy between several machine learning, deep learning architectures, and classic linear models. With a filtered sample of 3,407 AAPL OTM options and an optimal set of features, the research trains multiple models, including machine learning and linear regressors. Gradient Boosting, after tuning hyperparameters, is the best-performing model with an R^2 score of 0.9959, significantly outperforming linear baselines. Explainability methods like SHAP and LIME reveal moneyness to be the strongest predictor. The findings indicate that OTM options alone are not only sufficient but also optimal for IV modeling in this case. The work contributes to the existing literature demonstrating the effectiveness of tree-based models and proposes interpretable measures to explain the tail-risk dynamics of equity options.

1 Introduction

Implied volatility (IV) is a central notion in today's financial markets, and it represents the market expectation of an asset's future price direction as implied by option prices. It is a crucial input for options pricing, risk management, and making trading decisions. The accurate modeling of the IV surface, a representation of the way implied volatility varies with strike prices and maturities, is pertinent to portfolio hedging, volatility arbitrage, and derivative pricing Christoffersen and Jacobs (2004).

Implied volatility has been traditionally modeled through econometric models such as the Black-Scholes model Black and Scholes (1973). Providing interpretable structures but often making restrictive assumptions, such as constant volatility or normality of returns. Most recently, machine learning and deep learning techniques have been explored for their potential to more adaptively capture the non-linear and complex nature of IV surfaces Fischer and Krauss (2018), Zhang et al. (2023).

One such relatively unexplored domain is the separate role played by out-of-the-money (OTM) options in modeling the implied volatility of certain stocks, like AAPL. OTM options, in particular puts, are also known to reflect investor sentiment and tail-risk expectations more sensitively than ATM or ITM options Bates (1991); Weng and Xie (2024). Despite this, most IV modeling studies focus on more comprehensive indices (S&P 500) or aggregated options data and give little consideration to the impact of the depth and structure of OTM options on predictive performance and model specification.

1.1 Research Question

To what extent can out-of-the-money (OTM) options be used to accurately model the implied volatility (IV) surface of a single stock (AAPL), and which machine learning approaches, compared to traditional linear models, offer superior predictive performance and interpretability?

1.2 Research Objectives

The objectives to be addressed in the research include:

- To collect and preprocess option chain data for AAPL, focusing on out-of-the-money (OTM) call and put options.
- To engineer explanatory features relevant for implied volatility modeling.
- To implement multiple modeling techniques for predicting implied volatility.
- To evaluate model performance using standard regression metrics.
- To assess the contribution of each feature to model predictions using explainable AI.

2 Related Work

Implied Volatility (IV) refers to "a metric that captures the market's view of the probability of future changes in each security's price" [1] This is a critical concept in trading that affects decision-making, risk evaluation, and transaction costs. As a gauge of market sentiment, IV helps traders predict significant price movements. In addition to making strategic decisions, traders can use IV analysis to assess the possible risk of an option position and modify their portfolios accordingly.

This section provides a critical overview of the significant literature across traditional models like Black-Scholes and GARCH, which offer interpretability and closed-form solutions but rely on assumptions, the foundational concepts, and modeling challenges of Implied Volatility, comparing traditional vs machine learning approaches.

2.1 Implied Volatility Overview and Modelling Challenges

Understanding the fundamental properties of implied volatility surfaces is essential for addressing the research question, particularly regarding how OTM put options contribute to volatility modeling accuracy for individual equities.Cont and da Fonseca (2002) studied empirical properties of the IV surface over time, documenting its typical shapes and temporal evolution, providing empirical validation of volatility skew and term structure effects. However, it does not propose a forecasting model. Christoffersen and Jacobs (2004) emphasizes that different loss functions, like pricing vs hedging error, significantly affect IV model performance, which is critical for real-world IV applications. However, the model comparisons are rooted in econometric techniques, ignoring deep learning alternatives.

[1] https://www.investopedia.com/terms/i/iv.asp

The practical construction of implied volatility (IV) surfaces is outlined by Gatheral (2006), highlighting issues such as skew, smile, and term structures. However, the lack of modern machine learning perspectives makes it less useful for computational forecasting strategies. For its part, Carr and Wu (2008) demonstrated that implied volatility embeds market expectations of risk, which can be modelled as a risk premium, highlighting the role of macro uncertainty and investor sentiment in shaping the IV surface. Conversely, it focused mainly on the decomposition of volatility rather than on modelling surface shape.

In a more recent research, Li et al. (2021) examined the IV of 50 ETF options in China, demonstrating that "implied volatilities for in-the-money options and out-of-the-money options are larger than those of at-the-money options". Additionally, Weng and Xie (2024) investigates how deep out-of-the-money (DOTM) options better capture market sentiment, and they establish that low-frequency sentiment is strongly related to DOTM implied volatility dynamics. Demonstrating that model characteristics with the capacity to capture OTM degree can enhance predictive performance.

2.2 Traditional Models vs Machine Learning

The evolution from traditional to machine learning approaches is crucial for understanding the modeling effectiveness required to answer the research question. Fischer Black, Myron Scholes, and Robert Merton presented the innovative Black-Scholes model in the early 1970s. This model offers a way to price European options. The concept permits dynamic replication options utilising delta hedging by presuming constant volatility and ongoing trade. This framework was expanded and added to the Merton model for credit risk assessment. Subsequently, Cox et al. (1979) created a discrete-time framework for the binomial option pricing model, and Heston (1993) included stochastic volatility in option pricing models, which improves realism over Black-Scholes. However, it requires calibration and still lacks flexibility to model real-world IV surface deformations, especially under extreme market conditions.

Although Bartunek and Chowdhury (1995) demonstrated where traditional econometric models hold value through a direct empirical comparison between implied volatility forecasts using the Black–Scholes inversion and GARCH model forecasts, it does not consider more sophisticated ML models because it focuses on time series rather than cross-sectional option data. Bollen and Whaley (2002) empirically demonstrates that net buying pressure and order flow have a significant effect on the level and slope of implied volatility functions, particularly for index puts and equity calls. However, the model is descriptive, and there is a void in the use of machine learning methods that can potentially predict IV dynamics.

These traditional approaches provide interpretable frameworks but may lack the flexibility needed to capture the complex patterns in OTM options for individual equity modeling, suggesting the need for more advanced methodologies.

2.3 Machine/Deep Learning for Volatility and Option Pricing

Machine learning approaches offer potential solutions for the complex modeling requirements of OTM put options. Machine and Deep learning algorithms, especially time series networks, have been employed to forecast stock prices and movements. Fischer and Krauss (2018) and Nelson et al. (2017) conducted studies that demonstrated LSTM models significantly outperform traditional models for stock price movements and LSTM's power in capturing long-term dependencies, which applies to IV modelling. While Chen and Zhang (2019) integrated Long Short-Term Memory (LSTM) with attention mechanisms, for capturing smile and term structure in IV surfaces. Zhang et al. (2023)proposes the integration of excluding static arbitrage into neural networks to model IV surfaces; however, the model complexity may reduce transparency and replicability. In addition, some studies have implemented more robust models, for example Kim et al. (2024) uses physics-informed priors in a transformer-based CNN model to predict IV surfaces.

Furthermore, some studies have combined hybrid econometric plus deep learning architectures for volatility forecasting, for example, Xu et al. (2024) and Ulu (2024) integrate GARCH into LSTM architecture for short term stock volatility forecast, showing significant accuracy improvements over either method alone. In addition, while Gadhi et al. (2024) developed an study using WGAN-GP for synthetic data augmentation, supporting advanced LSTM/BLSTM performance in unstable periods, the study focuses on realized volatility remaining efficacy for IV surfaces unexplored. Focusing on the economic value of forecasting with ML, various authors have conducted implied volatility surface prediction using machine learning. Krauss et al. (2017) compares ensemble ML methods with DL in financial markets, which is relevant for benchmarking.

Given its crucial significance in option pricing, risk management, and market sentiment monitoring, modelling and forecasting implied volatility has been a major focus of the financial econometrics literature. Wiedemann et al. (2025) employ neural operators to interpolate IV surfaces from raw S&P500 options data with arbitrage-free securities while Wang et al. (2021) combines deep learning with financial mathematics ensuring mathematical consistency between the learned option prices and volatility surfaces. For its part, Liu et al. (2020) use a data-driven machine learning technique, artificial neural network (ANN), to calculate the implied volatility of American and European options. Moreover, Borochin and Zhao (2025) examines whether machine learning (ML) models can forecast implied volatility (IV) more accurately than traditional linear models using the same input variables. While these studies demonstrate the potential of machine learning for IV modeling, most focus on broad market indices rather than single equities, and few specifically examine the effectiveness of OTM options in their modeling frameworks.

2.4 Explainability and Ethical Considerations

Understanding model interpretability is important for evaluating the effectiveness of any approach to modeling IV with OTM options, particularly for practical implementation in single equity contexts. As with most neural network approaches, the models used lacks interpretability making difficult to understand the economic reasoning behind the implied volatility calculations, which may pose challenges for model transparency. Ribeiro et al. (2016) introduces Local Interpretable Model-Agnostic Explanations (LIME), a model-agnostic explanation tool widely used in financial ML to enhance transparency. Pimentel et al. (2025) Introduces an LSTM-based model for pricing European call options (S&P500 index) and reveals which input feature drive model predictions by implementing explainable AI using SHapley Additive exPlanations (SHAP). In addition, Shen (2023) builds a self-attention GRU model for pricing American-style options (SPY ETF) and applies SHAP to understand the influence of features on price predictions.

These explainability frameworks are relevant for understanding how OTM options contribute to IV modeling effectiveness, though their application to single equity contexts remains limited.

2.5 Gaps & Limitations

The literature formulates several axes of key results for our research question. Early models like Black-Scholes provide interpretable models of option pricing but rely on the unrealistic assumption of constant volatility, which lowers their usefulness at capturing the rich dynamics of implied volatility surfaces. GARCH and stochastic volatility models are improvements but remain reliant on restrictive assumptions. Machine learning techniques, in particular LSTM networks and attention mechanisms, have been demonstrated to excel at capturing non-linear relationships in volatility forecasting and can handle the complex temporal dependencies in IV surfaces. Hybrid models that combine econometric principles and deep learning are auspicious for improving accuracy without compromising some interpretability.

Despite these advancements, there are several crucial gaps that have a direct impact on the research question. Most studies focus nearly exclusively on call options or overall market indices with limited direct attention to put options or single equity application. There is no systematic study in the literature of how OTM options specifically contribute towards IV model performance, particularly for individual stocks rather than market indices. There is a lack of research on comparative effectiveness of option types (calls vs puts) and moneyness levels (ITM vs ATM vs OTM) for single equity IV modeling. Furthermore, the majority of machine learning uses in IV modeling do not have end-to-end explainability frameworks to understand why option types or moneyness levels are relatively more effective.

3 Research Methodology

3.1 Research Approach and Justification

This research adopts an empirical data-driven approach, combined with descriptive analysis, visualization techniques, regression modelling, and supervised learning techniques to capture the relationship between key features such as moneyness, day to expiration, spread, and implied volatility. The decision to focus exclusively on OTM options is theoretically motivated by several factors. First, and according to Bates (1991), Out-of-the-money (OTM) put options contain the most accurate information about extreme downside risk, as they are more sensitive to large negative price movements. Second, OTM call options disclose information about up volatility expectations free from the distorting effects of intrinsic value. Third, excluding at-the-money (ATM) and in-the-money (ITM) options, analysis discriminates the volatility smile effects from the mechanical interdependencies between intrinsic value and option prices.

Tree-based and boosting models such as Decision Tree, Random Forest, Gradient Boosting, and XGBoost perform best for static/tabular IV prediction, especially with noisy, nonlinear features common in OTM options. Neural networks such as MLP, LSTM, GRU, and 1D CNN can handle time-changing or very nonlinear IV surfaces, but require more data and fine-tuning. However, Linear models such as Linear, Lasso, Ridge, and Huber Regressor are interpretable and sufficient as baselines, but usually not sufficient for OTM implied volatility's complex shape modeling.

3.2 Data Collection and Processing

This research employs AAPL options chains obtained from Yahoo Finance [2] and covers a representative sample period of four days 10th, 11th, 12th, and 13th June) to capture diverse market conditions. However, after analysing the IV surface for each day, the final dataset excludes data from 12th June, which could affect the data and the development of the models. Before developing the models, an exploratory data analysis (EDA) is conducted to identify the characteristics, distribution, and the relationship between the variables and the target. Employing descriptive statistics, visualizations, and a correlation heatmap. The data preprocessing is implemented by converting relevant fields to numeric, filtering the dataframe to remove options that have not been traded, keeping open interest and IV greater than zero to ensure that the analysis is based on options that are actively traded and have meaningful data.

3.2.1 Feature Enineering

Core features are calculated, including moneyness, defined as the ratio of the underlying stock's spot price to the option's strike price, and bid-ask spread, which serves as a proxy for option liquidity. In addition to directly addressing the research question on modeling implied volatility using out-of-the-money (OTM) options, the methodology introduces a novel continuous variable, referred to as 'OTM distance'. This feature quantifies how far each option lies from being at-the-money (ATM), where the strike price equals the spot price, and is calculated as $(K-S)/S$, where K is the strike price and S is the spot price. The directionality of the measure, with positive values for calls and negative values for puts, allows the model to distinguish between OTM positions on either side of the volatility skew.

While moneyness is a well-established input in volatility surface modeling Fengler (2005), OTM distance provides additional granularity specific to OTM options by explicitly capturing the depth of the OTM position. This is supported by Fu et al. (2016), who show that IV asymmetry, particularly the elevated IV of OTM puts relative to ATM calls, is a significant predictor of cross-sectional stock returns and reflects non-linear pricing structures. Furthermore, empirical evidence indicates that IV tends to increase with distance from the ATM level, especially in the tails Li et al. (2021);Zhang et al. (2023).

Recent literature emphasizes the predictive power of fine-grained moneyness features in volatility surface modeling. Specifically, Luong and Dokuchaev (2018) demonstrates that Random Forests models outperform traditional linear methods in forecasting realised and IV by leveraging continuous, normalized predictors. In this study, OTM distance serves as a directional, normative feature that improves non-linear model learning and interpretability, and SHAP/LIME analyses confirm its high importance in predicting IV.

4 Design Specification

This part explains the approaches, frameworks, and architecture utilized for developing the suggested solution for IV modeling with OTM options of AAPL. Machine learning techniques and neural network-based models were employed to evaluate the predictive power of features engineered from options data. To predict the IV, a set of descriptive variables extracted from the options chain is used with feature selection that includes

[2] https://finance.yahoo.com/quote/AAPL/options/

'spread', 'moneyness', 'Days to Expiration', 'volume', 'open Interest', and a custom feature called 'OTM distance'. The final dataset contains 3.407 entries, which are split into a 20% test size with random state = 42, which is then scaled using StandardScaler.

4.1 Model Architecture Specifications

4.1.1 Linear Regression Models

- Linear Regression Using Statsmodels: This OLS model serves as a baseline model to offer coefficient interpretability and statistical significance testing for predicting IV through its specification:

  $$\mathrm{IV} = \beta_0+\beta_1\cdot\mathrm{Spread}+\beta_2\cdot\mathrm{Moneyness}+\beta_3\cdot\mathrm{DTE}+\beta_4\cdot\mathrm{Volume}+\beta_5\cdot\mathrm{OI}+\beta_6\cdot\mathrm{OTM}_{\mathrm{distance}}+\varepsilon \tag{1}$$

  The model assumes a linear relationship between option characteristics and the IV, serving as a reference for more advanced architectures.

- Regularized Linear Models: linear models like Ridge and Lasso regression are used to improve the linear model to address possible high correlation between features.

4.1.2 Ensemble and Non-Parametric Models [3]

- Random Forest: Applies bootstrap aggregating with decision trees, resulting in natural measures of feature importance and non-linear treatment of relationships by averaging a series of decision trees that have been trained on different subsets of data through ensemble averaging.

- Gradient Boosting: Constructs models sequentially and seeks to reduce prediction errors of previous iterations by constructing successive new models to predict the residual of the ensemble built so far.

- XGBoost: Regularized gradient boosting with advanced regularization techniques, efficient computation based on approximate tree learning, and intrinsic missing value handling.

- Decision Tree: Single tree-based model that recursively partitions the feature space based on feature values to reduce prediction error.

- Support Vector Regressor (SVR): Non-parametric method that maps input features to a high-dimensional feature space using kernel functions (RBF kernel in this pipeline) to determine non-linear associations. The model finds the optimal hyperplane that reduces prediction error within a tolerance region.

- K-Nearest Neighbors (KNN) Regressor: Instance learning algorithm that approximates IV as a function of the mean of k most similar neighbors in feature space. The model assumes that option properties with comparable values generate comparable IV.

- Huber Regressor: Brings the best of both L1 and L2 regularization together. It uses squared loss for small residuals and absolute loss for large residuals.

[3] https://scikit-learn.org/stable/api/index.html

4.1.3 Neural Network-Based Models [4]

- Multi-Layer Perceptron (MLP): Dense neural network with multiple hidden layers that can learn complicated non-linear relations.
- Long Short-Term Memory (LSTM): Recurrent neural network architecture that can learn sequential dependencies, though adapted for cross-sectional option data.
- Gated Recurrent Unit (GRU): A Lightweight version of LSTM with fewer parameters and which can counter overfitting.
- Convolutional Neural Network (CNN): 1D convolution layers intended to capture the local patterns of feature interactions.

4.2 Hyperparameter Optimization

The study uses a state-of-the-art hyperparameter optimization library, Optuna, to systematically search for the optimal model configurations. In this case, the best 3 ensemble models (XGBoost, Random Forest, Gradient Boosting) and the best 3 neural network (CNN, MLP, LSTM) models are optimized. The Optuna parameterization includes using Tree-structured Parzen Estimator (TPE) for efficient search, implementing early stopping to prevent overfitting, conducting 20 trials per model for comprehensive exploration and a different set of hyperparameters, and maximizing the R^2 score. Obtaining the Best Params: *'n estimators': 162, 'learning rate': 0.12573638174501423, 'max depth': 13, 'min samples split': 8, 'min samples leaf': 3, 'subsample': 0.8720802617662734, 'max features': None*

4.3 Evaluation Metrics

The performance of all implemented models is evaluated through metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). An informed decision about which model is the best at predicting IV is made by evaluating the lowest MAE, MSE, and RMSE with the Highest R^2 Score. This evaluation is discussed in section 6.

5 Implementation

The implementation is conducted using Python and the development environment Jupyter Notebook within Google Colab, providing cloud-based computational resources with T4 GPU acceleration for neural network training.

5.1 Data Transformation

The initial dataset contains 8,411 options across multiple expiration dates, used to plot the IV surface for each day. June 12 data is removed due to anomalies identified in IV, reducing the dataset to 6,297 options. Implementing the definition of the OTM condition, to separate 3,441 options where call options had strikes above the underlying price and put options had strikes below the underlying price. Finally, apply filters to

[4] https://www.tensorflow.org/guide/keras

Category	Libraries
Data Processing	Pandas, Numpy
Machine Learning	Scikit-learn, XGBoost
Deep Learning	TensorFlow with Keras API
Statistical Analysis	Statsmodels
Hyperparameter Optimization	Optuna
Model Interpretability	SHAP, LIME
Visualization	Matplotlib, Seaborn, Plotly

Table 1: Core libraries and frameworks

remove options with zero volume, zero open interest, or missing IV data. Resulting in **3,407** options suitable for model development.

5.2 Explanatory Data Analysis (EDA)

For every variable in the dataset, descriptive statistics are provided by the inferential analysis in Table 2, revealing that the moneyness distribution confirms a focus on OTM, with extreme outliers representing deep OTM options due to its mean being greater than 1 (1.73). The distribution of IV (mean = 46.8%) with a right skew (median ¡ mean), indicates the presence of some extremely high-volatility options. While the volume and open interest are strongly right-skewed distributions with high volatility, suggesting concentration of trading in certain contracts. In addition, the mean distance of OTM options from the strike is 0.40, with a moderate spread, showing that a typical OTM option is about 40% away from being at-the-money.

Statistic	Spread	Moneyness	Days to Expiration	Volume	Open Interest	OTM_Distance	Implied Volatility
count	3407	3407	3407	3407	3407	3407	3407
mean	0.3541	1.7350	275.66	670.63	3994.52	0.3981	0.4688
std	0.6517	3.6096	275.28	4345.97	7337.92	0.2818	0.4652
min	0.00	0.4361	0.00	1.00	1.00	0.0013	0.1152
25%	0.0186	0.7255	38.00	2.00	240.00	0.1616	0.2935
50%	0.0425	0.9891	189.00	12.00	1317.00	0.3502	0.3333
75%	0.2500	1.4749	464.00	105.50	4442.50	0.5795	0.4671
max	2.00	40.5540	920.00	98905.00	91764.00	1.2928	7.7500

Table 2: Descriptive Statistics of Model Features

Moreover, the correlation heatmap in the figure. 1 reveals (i) strong positive correlations between IV and moneyness (+0.49), which means deeper OTM options demand higher volatility premiums, and IV and spread (+0.6), which means wider spreads are associated with greater uncertainty and volatility. (ii) moderate correlation of OTM distance with DTE (+0.27) and spread (+0.29), which means options with the furthest expiration dates are usually further OTM, and options further from the money tend to have wider spreads. (iii) negative correlation of DTE with open interest (-0.26) and spread (-0.38), suggesting that longer-dated options have narrower spreads and tend to have lower open interest.

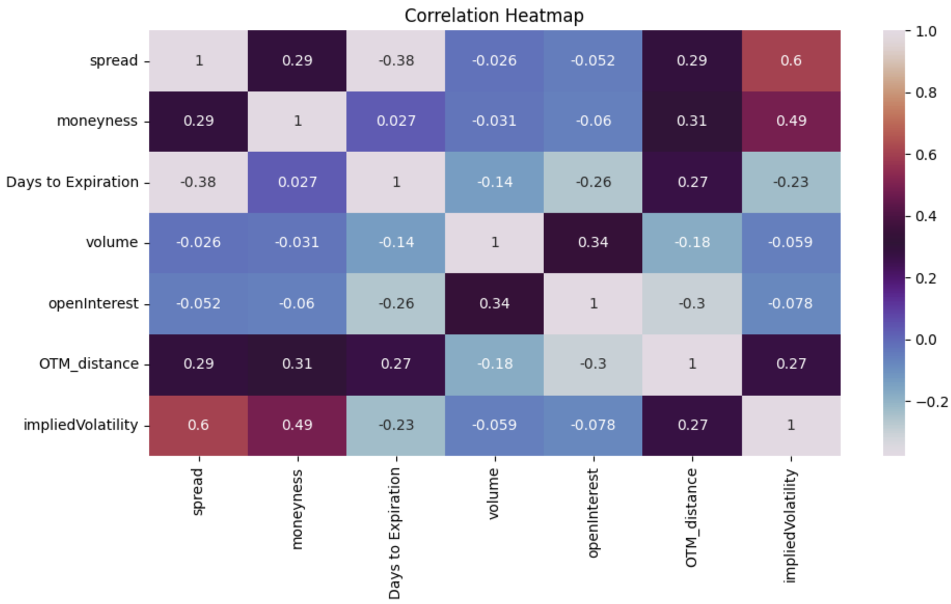


Figure 1: Feature Correlation Heatmap

5.3 Explainable (XAI) AI tools and Residual Analysis

Implementing Shapley Additive Explanations (SHAP) framework for the best model performance to produce feature importance rankings and summary plots to obtain an individual prediction explanation. The analysis quantifies the contribution of each feature to the model's predictions across the entire dataset. In addition to Local Interpretable Model-agnostic Explanations (LIME), to generate local model explanations that provide instance-specific feature importance assessments and prediction confidence intervals. Additionally, a comprehensive analysis of the residuals is generated using residual distribution histograms, Q-Q plots for normality assessment, and scatter plots of actual versus predicted values for bias detection.

6 Evaluation

In this section, the research question is addressed through statistical analysis and comprehensive model comparison, answering it through empirical evidence that demonstrates that OTM-only options contain sufficient information to achieve exceptional predictive accuracy ($R^2 = 99.59\%$) when combined with appropriate machine learning techniques.

The evaluation reveals that tree-based ensemble learning algorithms, specifically Gradient Boosting, achieve exceptional performance with R^2 (0.9959), over 38% points superior to that of standard econometric practice. The OTM distance feature has a consistent positive impact on predicted IV (0.10) while providing an interpretable measure of tail risk. Implied volatility estimation in OTM options is dominated by non-linear dependencies, 58.6% of predictive power being captured by moneyness effects through complex interactions rather than linearity.

6.1 Comprehensive Model Comparison

Table 3 shows the top 5 best models' performance. Gradient Boosting emerges as the best performing after optimization, demonstrating superior ability to capture complex non-linear relationships in options data.

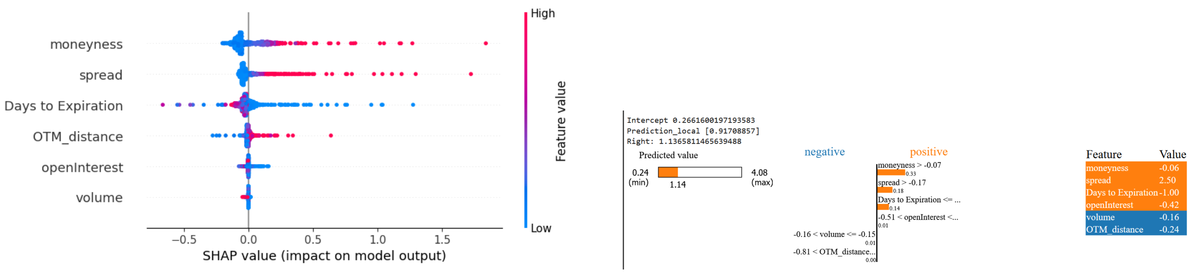
Rank	Model	R^2 Score	MAE	RMSE	Performance Category
1	Gradient Boosting (Optimized)	0.9959	0.0250	0.0489	Exceptional
2	Random Forest (Optimized)	0.9840	0.0181	0.0520	Excellent
3	LSTM (Optimized)	0.9476	0.0453	0.1133	Very Good
4	MLP	0.9450	0.0469	0.1036	Very Good
5	Lasso	-0.0005	0.2070	0.4424	Poor

Table 3: Model Performance Ranking

- The Gradient Boosting model, after optimization, indicates the model explains 99.59% of the variance in IV. MAE (0.0250) and RMSE (0.0489) show high accuracy and minimal prediction error.
- The Random Forest, after optimization, still offers solid results with R^2 98.40%. The higher MAE (1.81%) indicates slightly less accurate predictions, but maintains great practical utility and resistance to overfitting.
- The LSTM, after optimization, achieves 94.76% variance explanation. While the 4.53% MAE suggests the model can capture complex feature interaction, the "black box" nature can lead to limitations in practical adoption compared to tree-based models.
- The MLP, performs very good, R^2 (0.9450); however, the MAE (0.0469) and RMSE (0.1036) are the highest among top-tier models.
- The Lasso Regression performs poorly, compared across all implemented models, with negative R^2 (-0.0006), MAE (0.2070), and RMSE (0.4424), indicating substantially higher errors, and likely underfitting due to its strong regularization.

6.2 Feature Importance Analysis

Based on the model performance results, Gradient Boosting (GB) is used as the primary benchmark model, conducting feature importance and residual analysis for this model.



(a) SHAP plot (b) LIME explanation

Figure 2: Gradient Boosting XAI

The SHAP summary plot in Figure 2 reveals how each feature contributes to the GB model's predictions. Moneyness (0.1131) as the dominant predictor suggests that deeper OTM options show higher IV. While DTE (0.0652), as a moderate predictor, indicates that short-term options increase volatility. The OTM distance (0.0173) has a low to moderate impact on the predictions; furthermore, as the strike price moves away from the market price (red), the IV increases slightly. Volume (0.0017) has minimal impact, showing that higher volume slightly reduces volatility.

On the other hand, the LIME explanation provides an analysis based on the model prediction context: intercept (26.61%) represents the average IV across all options, the wide prediction range (0.24 to 4.08) indicates the model encounters substantial variability in the target variable across different market conditions. Similar to SHAP, Moneyness > 0.07 (+0.33) is the strongest positive contributor. For the AAPL option, the model predicted a relatively high implied volatility of 1.14, mainly because the option has a high moneyness and a wide bid-ask spread, both of which are strong indicators of higher volatility. Additionally, its short time until expiration further increased the predicted volatility.

6.3 Gradient Boosting Residuals Distribution

The residual distribution for GB is represented in Figure 3, which shows the distribution of prediction errors from the GB model. The residuals are tightly distributed around zero, and most errors fall within an approximate range of ±0.025, indicating that the model will tend to make quite small prediction errors on average. There is a mild positive asymmetry: the red dashed line is slightly to the right of zero, suggesting there is a slight tendency on average for the model to underestimate, but the bias is minor. A bell-shaped residual distribution is desirable since it would indicate that model errors are random rather than systematic. In this case, the model has been able to capture the underlying trends in data. There are very few residuals in the tails (outside ±0.05), indicating the model generates large prediction errors infrequently. This residual plot is characteristic of a Gradient Boosting model with good performance, good predictive ability, and low bias.

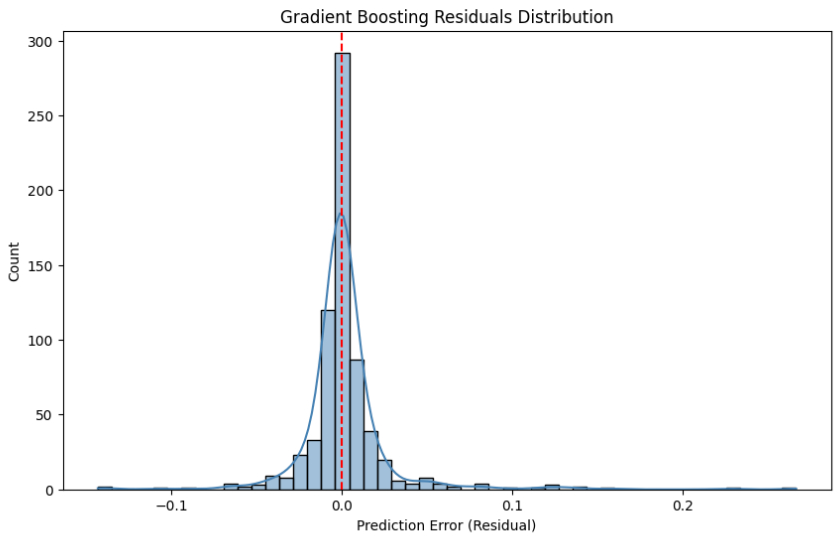


Figure 3: GB Residuals Distribution

6.4 Gradient Boosting Actual vs Predicted

Figure 4 performs the GB model through the comparison of actual and predicted values. The points are very heavily concentrated along the red dashed line (perfect prediction line), indicating that the model is very well capable of accurate predictions for all ranges of values. Model performance is acceptable in all ranges of true values (from the lowest to the highest) with no noticeable areas where the performance decreases significantly. Points are scattered adequately above the diagonal line with no noticeable tendency of over- or under-prediction, as indicated by the low bias in the residual histogram.

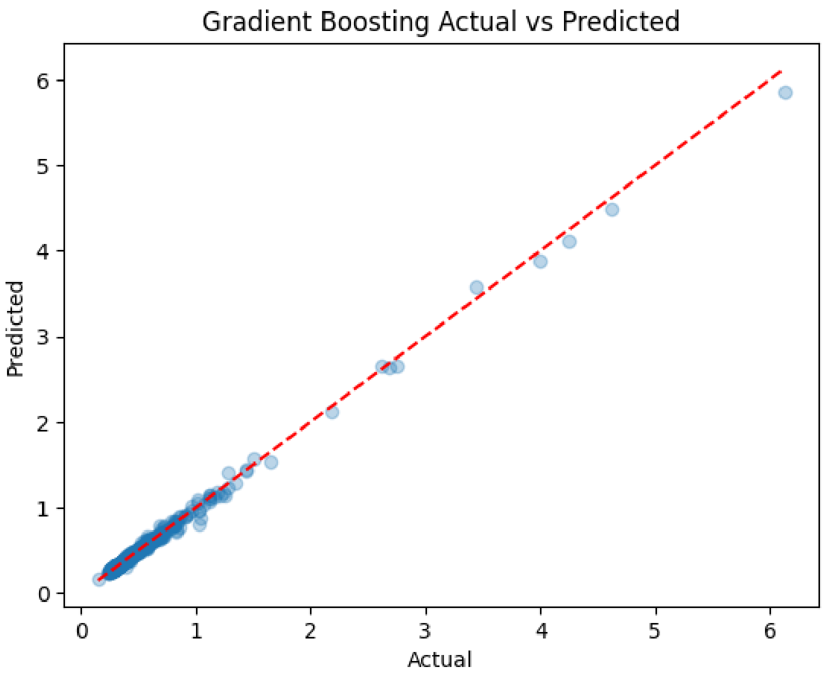


Figure 4: Actual vs Predicted

6.5 Q-Q Plot of Residuals

The Q-Q (quantile-quantile) plot in figure 5 graphs the distribution of the residuals of the Gradient Boosting model against a perfect normal distribution. Most of the points follow the red line of reference closely, especially in the middle part (approximately between theoretical quantiles -1 and +2), indicating that the residuals are approximately normally distributed. Some outliers are visible at both extremes; these heavy tails mean that occasionally the model makes greater prediction errors (positive and negative) than would be observed if the errors were perfectly normally distributed.

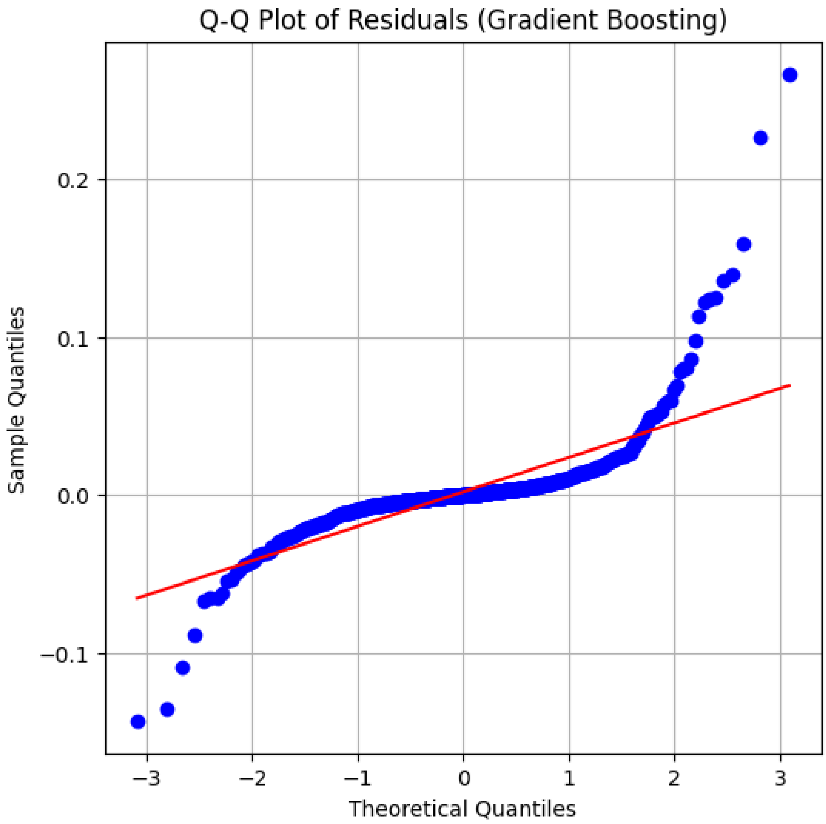


Figure 5: Q-Q Plot for GB

6.6 Discussion

This research explored whether OTM options contain sufficient information for accurate IV modeling of AAPL options. The optimized Gradient Boosting model performance R^2 (0.9959) demonstrates that OTM options alone contain sufficient information for highly accurate IV prediction. However, the highest R^2 scores across multiple models raise concerns about model generalization. The limited temporal scope (only 3 days of data) may have allowed the models to memorize instead of learning generalizable patterns.

The SHAP analysis revealed that moneyness was the most significant characteristic that influenced model predictions (SHAP importance: 0.1131), aligning with Weng and Xie (2024), following theoretical expectations that more in-the-money options pay higher volatility premiums. The custom OTM distance feature has a moderate contribution to model performance, validating the research hypothesis that quantifying distance from at-the-money strikes provides valuable predictive information. These findings support Li et al. (2021).

However, the research contradicts Bates (1991). While the current study supports the informational value of OTM options in general, it did not specifically validate the superior predictive power of put options compared to call options. The results align closely with Fischer and Krauss (2018), Krauss et al. (2017), and Chen and Zhang (2019), who established the superiority of deep learning approaches for capturing complex volatility patterns. The research presents some important limitations regarding the time frame, which limits the generalizability of the findings and fails to capture diverse market regimes that would test the robustness of the model. contradicting Cont and da Fonseca

(2002), who emphasizes the importance of time-evolving IV surfaces. The single option focus restricts the broader applicability of findings. The data collection period could lead to models that perform well under certain conditions failing during market stress.

Moreover, the comparison of linear, ensemble, and neural network approaches provides convincing evidence in favor of tree-based approaches as the most suitable solution to this specific problem. The creation of the continuous "OTM distance" feature is a useful, operationalizing theoretical concepts of volatility smile impacts into something quantifiable, offering a continuous measure of tail-risk positioning. The application of SHAP and LIME analyses addresses the interpretability gap of much machine learning work in finance, partially answering concerns in the literature about model transparency.

The results strongly support Fischer and Krauss (2018), Nelson et al. (2017), and Chen and Zhang (2019) regarding the superiority of machine learning techniques for financial prediction tasks. In addition, the dominance of tree-based models supports Zhang et al. (2023) and Kim et al. (2024) in highlighting the importance of capturing non-linear patterns in IV surfaces. There is a conflict between accuracy and interpretability since, although Gatheral (2006) emphasized interpretability in volatility modelling, recent research indicates that sophisticated machine learning models could be required for optimal performance. Furthermore, implementing some design improvements could guarantee better results, such as collecting data covering several years to capture various market regimes, implementing time-series cross-validation techniques to avoid anticipation bias, and incorporating VIX, interest rates, and economic indicators as suggested by Carr and Wu (2008).

7 Conclusion and Future Work

This research aims to investigate the predictive power of out-of-the-money (OTM) options in modeling the implied volatility (IV) surface of a single stock, Apple Inc. (AAPL). Specifically, the study aimed to determine whether OTM call and put options alone contain sufficient information to model the IV surface accurately and whether machine learning and deep learning models outperform traditional linear regression approaches. The research question was to assess the predictive capability as well as interpretability of such models, a consideration increasingly relevant in financial modeling with the growing demand for transparency and explainability. To achieve this research question, the project proposed six objectives, such as collecting and preprocessing AAPL options chain data, feature engineering, and developing a range of models from classical linear regressors to ensemble and deep learning models. The models were evaluated using the application of standard regression metrics such as MAE, MSE, RMSE, and R^2, while SHAP and LIME were utilized to provide interpretability to the predictions.

The research accomplished all its stated objectives. A 3,407 OTM options dataset was created through precise filtering and feature engineering. Models were implemented and assessed using a well-structured evaluation pipeline. The results showed that machine learning models significantly outperformed linear models in predicting implied volatility. Optimized Gradient Boosting was the best model, with an R^2 of 0.9959, suggesting very high predictive power. In contrast, traditional linear models such as Lasso and Ridge performed worse, demonstrating the non-linear and complex nature of the IV surface. The principal findings of this research illustrate that OTM options alone are not just sufficient for accurate IV modeling but can also yield a superior quality by avoiding the confounding effect of intrinsic value inherent in ATM or ITM options. The results also demonstrate that tree-based machine learning models are well-suited to capture the non-linear relationships between option features and implied volatility. The use of SHAP and LIME facilitated the explanation of model behavior in an interpretable way, revealing that moneyness and spread features were the most influential predictors across models. These observations can have important implications for traders, risk managers, and quantitative modelers who are looking to enhance the accuracy and interpretability of volatility forecasting systems. Nevertheless, these contributions, there are some limitations that need to be noted.

The data covered just three days of market data, which limited the temporal generalizability of the results. Additionally, the focus on a single stock, while providing the potential for clarity and control, restricts the scope of inference to other equities or market regimes. The model-building process was also cross-sectional and static, and therefore did not capture time-series dynamics of implied volatility. Furthermore, although neural networks were included, they did not outperform tree-based models in this static setting, which is possibly due either to architectural constraints or to the absence of sequential information. Future work can productively extend this research in several directions, such as application of the analysis to broader sets of equities and for longer time horizons to test for generalizability and robustness of the findings. Integration with temporal modelling using attention-based architectures to forecast the evolution of the IV surface over time. Macroeconomic factors can also be incorporated to assess how model performance differs under different market regimes. Furthermore, future work can investigate hybrid models that combine financial theory with machine learning architectures to extend realism along with accuracy. In summary, this research has demonstrated that OTM options are very valuable tools for modeling implied volatility on a single-stock basis and that machine learning models, in particular ensemble-based algorithms, provide both greater accuracy and interpretable insights relative to traditional econometric approaches. These findings contribute important insights to the financial ML literature and provide numerous interesting avenues for future practitioner and academic research.

References

Bartunek, K. S. and Chowdhury, M. (1995). Implied volatility vs. garch: A comparison of forecasts, *Managerial Finance* **21**(10): 59–73.

Bates, D. S. (1991). The crash of '87: Was it expected? the evidence from options markets, *The Journal of Finance* **46**(3): 1009–1044.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**(3): 637–654.
URL: *http://www.jstor.org/stable/1831029*

Bollen, N. P. B. and Whaley, R. E. (2002). Does net buying pressure affect the shape of implied volatility functions?

Borochin, P. and Zhao, Y. (2025). The economic value of equity implied volatility forecasting with machine learning, *Journal of Empirical Finance* **82**: 101618.

Carr, P. and Wu, L. (2008). Variance risk premiums, *The Review of Financial Studies* **22**(3): 1311–1341.

Chen, S. and Zhang, Z. (2019). Forecasting implied volatility smile surface via deep learning and attention mechanism.

Christoffersen, P. and Jacobs, K. (2004). The importance of the loss function in option valuation, *Journal of Financial Economics* **72**(2): 291–318.

Cont, R. and da Fonseca, J. (2002). Dynamics of implied volatility surfaces, *Quantitative Finance* **2**(1): 45.

Cox, J. C., Ross, S. A. and Rubinstein, M. (1979). Option pricing: A simplified approach, *Journal of Financial Economics* **7**(3): 229–263.

Fengler, M. R. (2005). *Semiparametric modeling of implied volatility*, Springer.

Fischer, T. and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions, *European Journal of Operational Research* **270**(2): 654–669.

Fu, X., Arisoy, Y. E., Shackleton, M. B. and Umutlu, M. (2016). Option-implied volatility measures and stock return predictability, *Journal of Derivatives* **24**(1): 58–78.

Gadhi, A. H. A., Peiris, S. and Allen, D. E. (2024). Improving volatility forecasting: A study through hybrid deep learning methods with wgan, *Journal of Risk and Financial Management* **17**(9).

Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*, John Wiley & Sons, Hoboken, NJ.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The review of financial studies* **6**(2): 327–343.

Kim, S., Yun, S.-B., Bae, H.-O., Lee, M. and Hong, Y. (2024). Physics-informed convolutional transformer for predicting volatility surface, *Quantitative Finance* **24**(2): 203–220.

Krauss, C., Do, X. A. and Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the sp 500, *European Journal of Operational Research* **259**(2): 689–702.

Li, P., Xian, A. and Lin, Y. (2021). What determines volatility smile in china?, *Economic Modelling* **96**: 326–335.

Liu, S., Leitao, , Borovykh, A. and Oosterlee, C. W. (2020). Machine learning to compute implied volatility from european/american options considering dividend yield, *Proceedings* **54**(1).

Luong, C. and Dokuchaev, N. (2018). Forecasting of realised volatility with the random forests algorithm, *Journal of Risk and Financial Management* **11**(4).
URL: *https://www.mdpi.com/1911-8074/11/4/61*

Nelson, D. M. Q., Pereira, A. C. M. and de Oliveira, R. A. (2017). Stock market's price movement prediction with lstm neural networks, pp. 1419–1426.

Pimentel, R., Risstad, M., Rogde, S., Rygg, E. S., Vinje, J., Westgaard, S. and Wu, C. (2025). Option pricing with deep learning: A long short-term memory approach, *Decisions in Economics and Finance* .

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, New York, NY, USA, p. 1135–1144.

Shen, Y. (2023). American option pricing using self-attention gru and shapley value interpretation.

Ulu, Y. (2024). Forecasting stock volatility via hybrid deep learning and garch family models: A case study from bist30, *Journal of Applied Mathematics and Computation* **8**(4): 280–285.

Wang, Z., Privault, N. and Guet, C. (2021). Deep self-consistent learning of local volatility.

Weng, J. and Xie, Y. (2024). Degree of irrationality: Sentiment and implied volatility surface.

Wiedemann, R., Jacquier, A. and Gonon, L. (2025). Operator deep smoothing for implied volatility.

Xu, Z., Liechty, J., Benthall, S., Skar-Gislinge, N. and McComb, C. (2024). Garch-informed neural networks for volatility prediction in financial markets.

Zhang, W., Li, L. and Zhang, G. (2023). A two-step framework for arbitrage-free prediction of the implied volatility surface, *Quantitative Finance* **23**(1): 21–34.