# What If Option Closing Prices Were Trustworthy? A Machine Learning Approach *

Alejandro Lopez-Lira    Mahendrajah Nimalendran    Matthew Son

First Version: March 2023
This version: July 2023

**Abstract**

Unlike stocks, options lack a robust mechanism to determine closing prices. Current practices relying on last-trade prices suffer from stale information, while closing quotes are unreliable due to consolidated trading and potential manipulation. We propose using machine learning to create a counterfactual auction price for the options market's close using the underlying stock prices. Our approach consistently outperforms traditional models, such as the Black-Scholes-Merton model, across all options series. The 4-PM mid-quotes and last-trade prices deviate from the counterfactual benchmark prices by 35% and 47%, respectively, suggesting significant efficiency gains can be achieved from implementing closing auctions.

Keywords: Options market, Closing auctions, Machine Learning, Big Data

*We thank Narasimhan Jegadeesh for helpful comments and suggestions.

# 1 Introduction

The daily closing prices of financial securities are arguably one of the most fundamental data points for researchers and practitioners. The NYSE and NASDAQ use a closing auction mechanism to provide reliable pricing and avoid last-minute consolidated trading and manipulative schemes. Equity options, however, do not have such a robust mechanism. With equity options available for thousands of underlying stocks with a range of moneyness and maturities, not all options possess sufficient transactions to guarantee effective price discovery at the close. However, options exchanges, funds, brokerages, and researchers use end-of-day mid-quotes or the last trade as a proxy for closing prices. We propose an easy-to-compute counterfactual benchmark using a machine-learning algorithm and the underlying stock closing auction price to determine reliable closing prices for options.

The Securities and Exchange Commission (SEC) does not have explicit regulations for reporting mutual funds' value of options holdings. Funds typically rely on the last transaction price for reporting and internal purposes. Brokerage firms report the last mid-quote (e.g., Robinhood) or the closing prices generated by the Options Clearing Corporation (OCC; e.g., Interactive Brokers).[1] Previous studies, such as Coval and Shumway (2001) and Muravyev and Ni (2020) use end-of-day mid-quote for daily option returns. However, there is limited discussion on the reliability of closing option prices and current practices, despite their importance in estimating NAVs for funds and margin requirements by brokerages.

Using the last transaction as a proxy for the closing price can have significant implications for infrequently traded options. This is because the last transaction for 61% of the options occur before 3 PM, while the underlying stocks continue to be actively traded until the closing auction. This can result in stale information, as the reported option prices may not reflect the latest price discovery. Our analysis reveals a significant trend where most options exhibit

[1] Options Clearing Corporation (OCC) develops a single price as of the close from multiple exchanges with a proprietary algorithm to ensure no-arbitrage conditions across strikes or time. It does not reflect the closing price as disseminated by any of its participant exchanges. https://ibkr.info/node/1199

infrequent trading activity as the market approaches its close. Specifically, approximately 80.4% of the options series on S&P 500 stocks have an average daily transaction count of less than ten. Moreover, only 16% of total transactions occur within the final 10 minutes of the trading hour. Among these transactions, 11% correspond to at-the-money (ATM) options, 4% to OTM options, and 1% to ITM options. These findings underscore the limited trading volume and a lack of liquidity observed in these options during the closing period.

Using the end-of-day mid-quote as a benchmark is also problematic for several reasons: (i) It is common to observe placeholder-like quotes that are unlikely to be traded at market close.It is unsurprising to see inconsistencies in the quotes because the market makers do not expect the quote to be fulfilled. (ii) Consolidated trading near the close can lead to large variability in price and returns. These factors can lead to noisy quotes at the close (Admati and Pfleiderer (1988) and Wood et al. (1985)). (iii) The widening bid-ask spread near close poses challenge to achieve price efficiency through arbitrage. Due to these factors, most options data vendors avoid using 4 PM quotes for calculating implied volatility as it produces an unreliable volatility surface.[2] Additionally, options generally have larger spreads, lower transaction frequency, and slower quote updates compared to the underlying stocks, especially for out-of-the-money (OTM) and in-the-money (ITM) options. The intraday transaction patterns and the quote behaviors near close shed light on the limitations and uncertainties of accurately determining closing prices for options, leaving us without a robust and reliable benchmark.

To address these issues, we propose a counterfactual benchmark for the closing option price based on the closing auction price of the underlying stock and a machine-learning (ML) model. We consider equity options as direct derivatives of the underlying securities (i.e., the price discovery is from the stock to the option) and assume that the correct spot option price is determined by the spot stock price at the time of the option transaction price. Our analysis includes a comprehensive evaluation of both American and European option pricing

[2]Daily implied volatility is calculated based on 3:45 PM or 3:59 PM option prices.

models and various machine-learning models.

Our findings reveal substantial deviations in both the mid-quotes and last transaction prices of the day when compared to our counterfactual ML benchmark, which is based on the stock close auction price. On average, we find deviations of 35% and 47% for the 4 PM mid-quote and last trading price, respectively. These deviations are particularly pronounced for smaller-sized firms and OTM options. Notably, the deviations of the last transaction price for ITM options are even more severe compared to the 4 PM mid-quote, with a difference of 11.57% versus 3.88%. This deviation holds economic significance as the current reporting practice of mutual fund holdings relies on historical prices, and ITM options possess greater asset value than at-the-money (ATM) and OTM options. We do not find significance in the impact of contemporary underlying volatility on the deviation. However, the occurrence of transactions throughout the day and the trade volume exert a significant influence in reducing the deviation.

In our study of options models for determining efficient closing prices, we consider widely accepted parametric models from Black and Scholes (1973) (BS or BSM) and Cox et al. (1979) (CRR).[3] For American options models, we also explore alternative estimation methods, including the finite-difference method with the Crank-Nicolson scheme (CN) and the Barone-Adesi and Whaley (1987) (BAW) model. Additionally, we explore machine learning algorithms: Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGB), Distributional Random Forest (DRF), Deep Learning (NN), and Elastic Net (EN). For simplicity and comparability with traditional models, we first restrict the inputs of the machine-learning models to the same variables utilized in conventional models. We adopt a standard random train-valid-test scheme. To enhance the predictive accuracy of the ML model, we then further augment the ML model inputs by including additional features, such as the VIX.

[3] Since the seminal work of Black and Scholes (1973) extended the models incorporate stochastic volatility, e.g., see Scott (1987); Hull and White (1987); Wiggins (1987); Heston (1993). Numerical methods (e.g., Broadie and Glasserman (1997); Longstaff and Schwartz (2001)) have been proposed for American options, which include premature exercise.

Our findings indicate that the input variables used in traditional options pricing models possess sufficient empirical information to explain the price variations across different options series. The traditional pricing models demonstrate a moderate level of accuracy, with out-of-sample $R^2$ values ranging from 71% to 82%.[4] In contrast, machine-learning models using the same input variables as the traditional models surpass their performance. Notably, the Gradient Boosting (GBM) model achieves an exceptional out-of-sample $R^2$ of 98.7%. This significant performance improvement can be attributed to the inherent flexibility of ML models, which allow for the incorporation of more complex functional forms and feature inter-dependencies when pricing options.

To further improve estimation convergence time and accuracy, we introduce additional features such as the moneyness $(S/X)$ ratio and VIX.[5] Unlike traditional option models, machine-learning models can easily incorporate new variables to improve the forecasts. Overall, the superior performance of ML models can be attributed to their ability to handle more complex relationships and dependencies among the variables. This flexibility enables ML models to achieve higher pricing accuracy than traditional models.

Often criticized for their 'black box' nature, machine-learning methods can offer insights when appropriately analyzed. Our study uses feature importance and regression analyses on these predictions to reveal the critical factors in different models. In the Black-Scholes-Merton model, volatility emerges as a crucial predictor of option prices, aligning with its foundational assumption of constant volatility. In contrast, the Gradient Boosting Machine model places greater importance on the time to maturity rather than volatility, indicating its effectiveness in capturing the temporal aspect of options pricing. By comparing feature importance across models, we gain insights into options pricing dynamics and understand how different models prioritize input variables.

[4]The CRR and CN algorithms are winsorized by 0.01% due to extreme predictions caused by non-convergence.

[5]Cao et al. (2020) find that the VIX index can produce a considerable improvement for ML models in the context of implied volatility. We use the VXX ETF as a proxy to match the most timely information at the 1-minute level.

The contemporaneous stock price serves as a critical input for our option pricing models, and we specifically utilize the closing auction prices. Extensive prior literature has examined the efficiency and benefits of closing auctions in the stock market, particularly during their initial implementations. For example, Pagano and Schwartz (2003) study the introduction of electronic call auctions at the market closing from Euronext Paris in 1996 and 1998. Their research findings indicate reduced execution costs and improved price discovery for the overall market due to the closing auctions. Barclay et al. (2008) argue that order consolidation within closing auctions leads to efficient prices based on their analysis of staggered auction implementations on NYSE and NASDAQ in 2004. Bogousslavsky and Muravyev (2022) conclude that the closing auction price is generally robust and efficient, with deviations being non-informational and quickly reverting. Hu and Murphy (2021) as well as Jegadeesh and Wu (2022) document price resiliency and differences between NYSE and NASDAQ call auctions at the close. Other researchers have also presented evidence supporting the benefits of auction markets. Budish et al. (2015) advocate for implementing frequent batch auctions over continuous trading due to the benefits of the auction mechanism. Similarly, Plante (2017) argue that order centralization would reduce transaction costs and enhance price efficiency in the corporate bond market. In summary, the recent studies provide empirical support for the efficiency, price discovery, and cost reduction benefits of closing auctions and auction mechanisms in various markets, reinforcing the use of closing auction stock prices as a critical input in our option pricing models.[6]

Machine learning models, particularly Neural Networks (NNs), have been used as a nonparametric method for option pricing since the early 1990s (Malliaris and Salchenberger (1993), Hutchinson et al. (1994) are some of the early work). Hutchinson et al. (1994) demonstrate

[6] Several studies have examined the mispricing of stocks and options at the market close, Bogousslavsky (2021) find that the end-of-day mispricing is pronounced for stocks with high overnight risk and unreliable quotes in the open. Muravyev and Pearson (2020) argue options effective spreads can have an upward bias because spot options quotes can be stale and asymmetric for fair option value. Barbon et al. (2021) discuss the intense trading activity at the close of the stock market triggered by options market makers' delta hedging.

the use of neural networks in enhancing the accuracy and computational efficiency of options pricing models. Ferguson and Green (2018) apply deep learning techniques to basket options, while Cao et al. (2020) find that incorporating the Volatility Index (VIX) improves model accuracy. Additionally, Andreou et al. (2023) find firm characteristics improves the model has predictive power. Ruf and Wang (2020) conduct a comprehensive literature survey on using machine learning, particularly artificial neural networks, in option pricing and hedging. They highlight the significance of volatility estimation choices in determining model outcomes and conclusions. The prior studies have mainly focused on very liquid index options (S&P500, EU STOXX 50), while our research is focused on less liquid individual stock options prices at the close of trading.

Despite the extensive research in pricing options, discourse on price efficiency and the enactment of call auctions at the close of the options market is seldom found in the literature and policy discussions. Our study contributes significantly to the existing literature by examining the textbook models and evaluating their applicability and relevance in the context of closing auctions. Additionally, we propose a novel and efficient empirical methodology that leverages machine learning to create a counterfactual auction price at the close of the options market. Further, we assess how these indicators deviate from the ML counterfactual price and observe notable disparities.

To the best of our knowledge, this study represents the first attempt to investigate the potential implications of closing auctions in equity options. By addressing this gap in the existing literature, we show auctions have the potential to enhance information efficiency in the options market at the close. Our research makes two significant contributions. First, we address the existing issues associated with the closing of options and highlight the need for improvement in current practices. To assess deviations accurately, we employ a state-of-the-art machine learning technique that demonstrates superior performance compared to traditional option pricing models. To the extent of our understanding, this study also represents the first attempt to investigate the information content of closing prices in the options market.

The second contribution is to the option pricing literature by exploring the implications of machine learning. Machine learning models possess substantial explanatory power, indicating that any discrepancies between theoretical predictions and empirical observations are primarily a result of the functional form used in traditional models rather than a deficiency in the information content contained within the input variables. We employ a model-agnostic interpretable machine learning approach with regression and feature importance analysis to illustrate the differences between conventional and machine learning models. This final technique aids us in understanding the underpinnings of these divergences, offering a more comprehensive perspective on option pricing practices.

The structure of this paper develops as follows: Section 2 presents the data and provides a descriptive analysis of the options market. Section 3 examines and contrasts standard parametric pricing models with machine-learning models, discussing their respective performance and interpretation. Section 4 presents the primary empirical findings of our study. Finally, Section 5 summarizes the conclusions drawn from our research.

## 2 Data and Descriptive Statistics

Our dataset is constructed from multiple sources, including the Chicago Board of Exchange (CBOE), OptionMetrics (OMetrics), Center for Research in Security Prices (CRSP), New York Stock Exchange Trades and Quotes (TAQ), and the U.S. Department of Treasury. Our analysis primarily relies on CBOE Trades and Quote data provided by the Options Price Reporting Authority (OPRA).

The OPRA dataset encompasses comprehensive historical option transactions from all U.S. exchanges, providing a high-frequency view with millisecond precision. It includes essential information such as the best quotes for both options and stocks at the time of trading and volume data. By leveraging this dataset, we gain insights into the dynamics of options trading.

Our study covers a sample period from October 2019 to March 2021, comprising a substantial volume of 1.7 billion transactions recorded across 370 business days. To ensure focus and relevance, we limit our analysis to equity options listed on the S&P 500 throughout the sample period. Non-standard options, including FLEX options, are excluded from the study. Additionally, we eliminate records with uninformative or likely erroneous data, such as those featuring equal or zero best bid-ask quotes and negative trading volume. For quote data, we utilize information recorded at a frequency of 1 minute, covering the entire year of 2020. The stock closing auction[7] and mid-quote prices at the open and close are obtained from the New York Stock Exchange Trades and Quotes (TAQ) data.

To calculate implied volatility and base Greeks (Delta, Gamma, Vega, Theta), we employ the widely used Black-Scholes-Merton (BSM) model. Dividend yields are derived from the CRSP dataset, while risk-free rates are approximated using month-level interpolated yields from Treasury data, matched to each option's expiration. Underlying stock volatility is calculated using 1-second stock returns over a 10-minute interval before the options transaction time, and the volatility estimate is then annualized. This provides a timely ex-ante estimate of volatility that is current in our enhanced ML model. We also include the VIX index (Cao et al. (2020) have shown that VIX significantly contributes to explaining the volatility surface).

Table 1 provides summary statistics for the key variables from the OPRA data that play a critical role in analyzing and benchmarking the performance of machine learning algorithms and traditional option pricing models. These variables include standard inputs used for the BSM model: strike price (K), trade price, mid-quote prices of the option (P) and the

[7] Specific trading condition variables denoted by "6" or "M" indicate the close, while "O" and "Q" represent the open auction price.

underlying stock (S), years to maturity ($\tau$), underlying volatility ($\sigma$), dividend yield ($q$), and the risk-free rate ($r$). Additionally, the study considers the base Greeks and implied volatility derived from the Black-Scholes-Merton (BSM) model.

[Insert Table 1 Here]

The data consists of more than 716 million transactions. The stock price and the strike price distribution are right-skewed due to the large stock prices for a few stocks during our sample period. Further, the distribution of options trading volume is right-skewed, with mean and median contract volumes (trade size) of 6.32 and 1.78, respectively, consistent with relatively low levels of trading activity for a significant portion of options contracts.

Figure 1 shows the fraction of trading activity by number of stocks. A significant amount of trading is concentrated around the top 15 stocks (AAPL options alone account for more than 10% of the volume), while the top 100 stocks account for more than 80% of the volume.

[Insert Figure 1 Here]

The average and the median transactions price per option are \$6.65 and \$2.35. It is observed that the first quartile and median values amount to \$0.68 and \$2.35 with a right-skewed distribution. These figures highlight the frequent occurrence of retail-sized options trading, where smaller-sized transactions are prevalent. It is notable that the median and standard deviation of transaction price slightly above of mid-quote. Furthermore, the average relative bid-ask spread, amounting to 11% per transaction, aligns with findings from other research studies. This observation suggests that the bid-ask spreads in the options market exhibit consistency in magnitude.

## 2.1 Options Intraday Trading

Regarding intraday trading, we continue to observe a U-shaped pattern with trading activity concentrated and more pronounced near the market open and close, similar to Chan et al.

(1995) and Bergsma et al. (2020). This U-shaped pattern indicates higher trading interest, liquidity, price discovery, and information efficiency during these periods.

[Insert Figure 2 Here]

Figure 2 illustrates the temporal distribution of the last transaction times for each trading day across various option series, with time intervals of 10 minutes. Notably, over 75% of the options have their final trade before 3:40 PM. Furthermore, many options show limited or no trading activity throughout the trading session, with a noticeable surge in trading volume observed during the market opening. Importantly, this trading pattern remains consistent across options of varying moneyness levels. These findings indicate that most options trading occurs earlier in the trading day, contributing to the challenge of obtaining reliable prices at the trading close.

## 2.2 Options Quoted Spread near Close

Figure 3 presents boxplots depicting the distribution of bid-ask spreads for options and the corresponding option-to-stock bid-ask spread ratios across different intraday trading times. The bid-ask spread is higher closer to the close and exhibits higher variability. Moreover, the effect is larger for OTM options. We see a similar pattern in the option-to-stock spread ratio.

[Insert Figure 3 Here]

There is a noticeable widening in the bid-ask spread of options starting from 10 minutes prior to the market close. This widening spread reflects heightened illiquidity, indicating an increase in the uncertainty of price determination during this period. We posit that the increased spread at the close is more closely linked to options market makers' inventory risk, rather than asymmetrical information risk when compared to the higher spread observed at market open (Bergsma et al. (2020) and Muravyev (2016)). This rise in illiquidity can reduce incentives for arbitrage and decrease price efficiency near close.

# 3 Machine Learning Models and Estimation

According to the "No Free Lunch" theorem in machine learning, there is no universally superior algorithm that outperforms others across all problem domains (Wolpert and Macready (1997)). The effectiveness of an algorithm is contingent upon the specific problem, the characteristics of the data, and the underlying context. Consequently, a comprehensive exploration and comparison of multiple algorithms are required to identify the most suitable model for a particular task. In our study, we meticulously evaluate the performance of the five most popular machine learning algorithms, employing robust hyperparameter tuning techniques. Additionally, we incorporate classical European and American option pricing models as benchmarks for comparison.

## 3.1 Structural Models

The Black-Scholes-Merton (BSM) European options model and the Cox-Ross-Rubinstein (CRR) binomial tree model have gained widespread adoption among researchers and practitioners, and we mainly use them as our benchmark reference point. Both American and European style option pricing models share common input variables, which can be concisely summarized by Equation 1:

$$p_t^y = f(S_t, K, \tau, r, \sigma, q), \tag{1}$$

where $p_t, S_t$ are the prices of the option and stock at time $t$, $\tau = T - t$ is time until maturity, and $y \in \{Put, Call\}$ specifies the type of the option. While most variables used in empirical estimations are often employed without further specification, the volatility measurement requires careful consideration. The BSM model assumes constant volatility over the option's maturity, and historical stock returns are used to estimate the volatility input for the models.[8]

[8]Alternatively, according to the literature survey conducted by Ruf and Wang (2020), researchers have used a wide range of volatility specifications, including calibrated volatility, GARCH-generated volatility,

We employ realized volatility calculated over the most recent 10-minute period as a proxy for volatility in all our models to provide timely information. As part of our robustness checks, we present more specifications for volatility estimates in the Appendix.

## 3.2 Machine Learning Models

We use the functional form described by Equation 2 to facilitate the formulation of our machine-learning algorithms.

$$p = f(X; \theta) \tag{2}$$

In the context of our estimation process, the trained model $f$ incorporates the feature matrix $X$ and a set of estimated parameters $\theta$ determined by the algorithm. The subsequent table provides a detailed account of our estimation process's overarching steps and procedures.

historical volatility, contract-specific volatility, historical-implied volatility, historical-at-the-money volatility, and volatility index among other methodologies.

**Algorithm 1** General description for estimating option price using Machine Learning

**Input:** A training dataset $\mathcal{D}$, a validation dataset $\mathcal{T}$, hyperparameter space $\mathcal{H}$
**Output:** A trained model $M$
**Initialize:** Randomly initialize hyperparameters $\in \mathcal{H}$
**Initialize:** Set the patience $p$
**for** $epoch = 1, \ldots, E$ **do**
  **for** $i = 1, \ldots, I$ **do**
    **Select a data point** $x_i \sim \mathcal{D} \in \{S, K, \sigma, \tau, r, q, y\}$
    **Predict the target** $\hat{y}_i = f(x_i; \theta)$
    **Calculate the loss** $L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell(y_i, \hat{y}_i)$
    **Update the parameters** $\theta \leftarrow \theta - \eta \nabla L(\theta)$
  **end for**
  **Calculate** $L(\theta)$ **on** $\mathcal{T}$: $L_{\text{val}}$
  **if** $L_{\text{val}} < L_{\text{best}}$ **then**
    **Set the best loss** $L_{\text{best}} = L_{\text{val}}$
    **Set the best parameters** $\theta_{\text{best}} = \theta$
  **else if** $L_{\text{val}} - L_{\text{best}} > p$ **then**
    **Stop training**
    **Return the best model** $M = (f, \theta_{\text{best}})$
  **end if**
**end for**
**Return the trained model** $M = (f, \theta_{\text{best}})$

Machine-learning algorithms can deal with an arbitrary number and type of input variables. Expanding the input variables and incorporating additional data can significantly enhance their performance. However, in the first exercise, we intentionally impose limitations on the input variables to be identical to the ones used by the classical models. First, this allows us to investigate scenarios where traditional options pricing models may fail due to either 1) insufficient information within the model or 2) limitations in the functional form. Moreover, this approach facilitates the application of interpretable machine learning techniques, enabling us to gain insights into the importance of individual variables, which holds significant implications for the options pricing literature.

In the asset pricing literature, employing a rolling or recursive sample-splitting scheme for forecasting is customary. This scheme involves shifting the training and validation samples forward to encompass more recent data, as exemplified by Gu et al. (2020). This approach is commonly employed to address potential violations of the underlying time series structure and information leakage. However, it comes at the cost of reduced computational efficiency, necessitating recursively re-estimating the model.

The fixed scheme is another commonly employed method for sample splitting, whereby the data is randomly partitioned without replacement according to a predetermined split ratio. However, caution must be exercised when utilizing this approach, mainly when the features encompass identifying variables and time-specific information, as it could potentially result in data leakage. Given the nature of option pricing models, which rely not on time-specific information but on the remaining lifespan of the options contract ($\tau$), the fixed-scheme approach is suitable for our study. By aligning our input variables with the structural models, we ensure that the machine-learning models do not incorporate any time-specific information, thus maintaining the validity of our analysis.

We adopt a two-fold approach to streamline the estimation process considering the multitude of hyperparameter dimensions and algorithms involved: random grid search method and train-valid-test split. This approach enables us to efficiently explore the hyperparameter space and evaluate the performance of various algorithms. In our implementation, we allocate 70% of the data for training purposes, while 15% is dedicated to the validation subset.[9] This validation set serves the crucial function of fine-tuning the hyperparameter search results within each algorithm. The remaining portion of the data is reserved exclusively for the final testing phase, ensuring an unbiased and independent assessment of the model's performance.

[9]To validate the robustness of our approach, we conducted a comparative analysis between the train-valid-test split and 5-fold cross-validation methods. Remarkably, both methods exhibited similar performance outcomes, while the former significantly reduced the overall estimation time.

## 3.3 Model Estimates and Performance Evaluation

### 3.3.1 BSM and CRR Model

In this section, we initiate our discourse by exploring price estimates utilizing conventional models as benchmarks, namely the Black-Scholes-Merton (BSM) model and the Cox-Ross-Rubinstein (CRR) model.

[Insert Table 2 Here]

Our empirical analysis uncovers the performance of the BSM model in estimating option prices, revealing a Mean Absolute Error (MAE) of 2.6 and a Root Mean Squared Error (RMSE) of 10.871. Notably, the BSM model exhibits an out-of-sample R-squared value of 81%. However, the substantial disparity between the MAE and RMSE values raises concerns regarding the distribution of prediction errors. The large discrepancy between the MAE and RMSE values indicates a non-uniform distribution of prediction errors. Specifically, a larger RMSE relative to the MAE suggests that the model is more sensitive to more extensive errors.

Interestingly, despite being designed for European-style options, the BSM model displays lower average bias when compared to American options models[10]. Alternative American options models, such as Cox-Ross-Rubinstein (CRR), Barone-Adesi-Whaley (BAW), and Crank-Nicholson (CN), yield MAE values around 2.7, RMSE values around 12, and R-squared values ranging from 70% to 77%.

### 3.3.2 Gradient Boosting Machine

In contrast, our analysis reveals that machine-learning models consistently outperform structural models. Notably, the gradient boosting machine (GBM) exhibits superior performance

[10] Nevertheless, it is essential to highlight that the evaluation of the Mean Percentage Absolute Error (MPAE), defined as $100 * \sum_{i=1}^{N} |\ln(\hat{y}_i/y_i)|/N$, favors the performance of American options models over the BSM model.

across all evaluation metrics, boasting a remarkably low MAE of 1.2 and RMSE of 2.9, surpassing both the BSM and CRR models. Our performance results highlight that structural models, including the BSM and CRR models, tend to exhibit more extensive estimation errors than the GBM model. It is worth noting that the overall out-of-sample R-squared value of 98.7% should be interpreted cautiously, as observations with large nominal prices can influence R-squared metrics based on a level variable.

Crucially, our estimation demonstrates the stability and reliability of our models by examining the consistency of performance metrics across the training, validation, and test samples. This outcome underscores the effectiveness of our hyperparameter tuning process in achieving a suitable balance between underfitting and overfitting. By attaining this balance, our models exhibit robustness in their predictive capabilities and provide confidence in the generalizability and validity of our findings.

The substantial explanatory power exhibited by the option price estimation is interesting, considering the stringent adherence to standard option pricing variables as our input. This significant outcome suggests that any disparities between theoretical predictions and empirical observations are primarily attributable to the functional form employed rather than a deficiency in the information content encapsulated within the input variables. Thus, our findings underscore the notion that the selected input variables possess ample information to account for the observed variations in equity options, reinforcing the robustness of our approach.

[Insert Table 3 Here]

Table 3 offers additional insights into the performance of various models based on the moneyness of options. The evaluation metric used in this analysis is the Mean Percentage Absolute Error (MPAE), which measures the relative error percentage. It is observed that all machine learning models exhibit lower errors for at-the-money options, followed by in-the-money and out-of-the-money options. Structural models outperform machine learning models in explaining ITM options. For instance, the Black-Scholes-Merton (BSM) and Cox-Ross-Rubinstein (CRR) models achieve Mean Absolute Error (MAE) values of 2.29 and 2.5, respectively, while the Gradient Boosting Machine (GBM) model reports a slightly higher MAE of 2.85. However, this result is reverted once we include additional explanatory variables.

Reducing bias in the model is crucial as it serves as a counterfactual benchmark price. Building upon these findings and the identified feature importances, which will be discussed in the following section, we explore adding additional input variables to enhance model performance. The study by Cao et al. (2020) highlights the improved accuracy achieved by incorporating a volatility index measure. In our first modification (GBM2), we include the S&P 500 VIX index quote and time-weighted volatility ($\sigma\sqrt{\tau}$) following the Black-Scholes-Merton (BSM) model. The underlying volatility ($\sigma$) is weighted by the square root of the time-to-maturity ($\sqrt{\tau}$), as illustrated in the BSM model equation. In our second modification (GBM3), we introduce the absolute difference ($S - K$) and relative moneyness ($S/K$) of the option instead of the weighted time-to-maturity. This modification aims to reduce the model estimation time while considering the strike price and stock price, which were identified as unanimously important features.

Including additional variables results in significant improvements across multiple performance metrics, including Mean Absolute Error (MAE), Mean Percentage Absolute Error (MPAE), Root Mean Squared Error (RMSE), and out-of-sample R-squared. In the first modification, the MAE on the test set decreased from 1.276 to 0.584 (a reduction of 54%). A further enhancement is achieved in the second modification, where the MAE reduces to 0.395 (a drop of 69%). Similarly, the RMSE demonstrates a notable reduction of 55% and 71% in the first and second modifications, respectively.

[Insert Table 4 Here]

Table 4 Panel A focuses on the performance evaluation of the Gradient Boosting Machine (GBM) models, including train, validation, and test outcomes based on different specifications. Our hyperparameter tuning avoids underfitting or overfitting, resulting in robust model performance. Notably, the GBM3 model consistently outperforms the other specifications across all performance metrics, demonstrating its superiority in predictive accuracy.

Moving to Panel B, we delve into the models' performance on the test dataset concerning options moneyness. Remarkably, the GBM3 model exhibits a substantial improvement for in-the-money options, reducing the average error from 39% to a mere 5%. This notable enhancement underscores the efficacy of the GBM3 model in capturing and predicting the nuances associated with different option scenarios.

[Insert Figure 6 Here]

Figure 6 depicts the relationship between actual option prices and the corresponding predictions generated by our four selected models. Distinct patterns emerge concerning the magnitude of prediction errors and variances exhibited by the Black-Scholes-Merton (BSM) and Cox-Ross-Rubinstein (CRR) models. Specifically, these structural models demonstrate more considerable prediction errors and larger variability as the price level decreases. However, it is noteworthy that these models exhibit relatively favorable performance for options with higher price levels, particularly those categorized as in-the-money. In contrast, the machine learning models, namely the Gradient Boosting Machine (GBM) and GBM3, display comparatively more minor variations in prediction accuracy across different price levels.

## 3.4 Interpreting Machine Learning Model

One of the prominent approaches for assessing interpretability in machine learning models involves examining feature importance, which remains agnostic to specific model architectures or algorithms.[11] Feature importance refers to assessing and quantifying the relative influence or contribution of input variables (features) in a predictive model's overall performance. It helps identify the most influential factors contributing to accurate predictions,

[11] The feature importance is typically based on the model's internal mechanisms, such as permutation importance, Gini importance, or coefficients derived from regularization such as Lasso or Ridge regression.

enabling researchers to validate existing theories or hypotheses. Furthermore, feature importance analysis aids in feature selection, where less informative or redundant features can be excluded, improving model efficiency and reducing overfitting.

We present a feature importance heatmap derived from the analysis using five distinct machine learning algorithms. The heatmap provides valuable insights into the relative significance of different variables in our models.

[Insert Figure 5 Here]

Figure 5 shows that all machine-learning algorithms unanimously identify the stock price ($S$) and the strike price ($K$) as the most influential variables for predicting option prices. This consensus across algorithms highlights the critical role played by these variables in determining option prices. Interestingly, it is worth noting that the importance of volatility appears to be relatively lower than conventionally believed. Furthermore, the best-performing algorithm (GBM) exhibits a moderate emphasis on the stock price ($S$) and strike price ($K$) in comparison to the other algorithms employed in the study. This observation suggests that GBM places less importance on these variables, although they remain crucial for accurate prediction.

As demonstrated by Hull and White (1987), the BSM model exhibits deviations from its predicted prices in situations involving stochastic volatility, mainly when there is a correlation between stock price and volatility. Notably, these deviations tend to increase as the option's time to maturity lengthens. The relatively diminished significance of volatility in our analysis can be attributed to adopting historical volatility ($\sigma$) as a proxy. Instead, we observe that the time to maturity ($T$) variable encompasses more relevant information about volatility for each option. This observation implies that the impact of volatility on option pricing may be better captured by considering the time remaining until the option's expiration. This finding highlights the potential for a more accurate representation of volatility's influence on option pricing.

To further enhance our understanding, we examine the feature importance in the Black-Scholes-Merton (BSM) model and its comparison to the Gradient Boosting Machine (GBM). By exploring the significance of different variables in the BSM model, we aim to elucidate their role in influencing option pricing. To facilitate a juxtaposition, we trained a separate GBM model to replicate the BSM model by utilizing the BSM estimated price as the target variable instead of the mid-quote option price.[12]

[Insert Figure 5 Here]

Figure 5 showcases the feature importance of the BSM model. Compared to Figure 5, both models identify the strike price (K), the stock price (S), and the time to maturity as the most significant variables in determining option pricing accuracy. Conversely, neither model attributes considerable importance to the dividend yield ratio as it does not contribute substantially to improving model performance. However, a notable disparity arises in treating underlying volatility ($\sigma$) and the risk-free rate ($r$). The GBM model places less emphasis on volatility, accounting for less than 10% of the variable importance. In contrast, the BSM model assigns greater importance to volatility, constituting more than 50% of the variable importance.

[Insert Table 6 Here]

We conducted a regression analysis based on the percentage error of model estimates compared to the mid-quote of option prices. Both models showed more extensive deviations for out-of-the-money options and less extensive deviations for in-the-money options than at-the-money options. Put options were generally predicted more accurately. Transaction size did not show any significant economic relationship with model accuracy. We observed a non-linear relationship between maturity and model accuracy, with options of moderate maturity being relatively better predicted than short and very long maturities. However, there are differences between the models. The Black-Scholes-Merton (BSM) model makes more

[12]GBM successfully replicated the BSM model with minimal errors.

accurate predictions for assets with higher volatility, while the GBM model tends to have slightly higher error rates in such cases. Furthermore, the BSM model performs relatively worse for equity options of large-cap firms, whereas the GBM model demonstrates superior performance for options associated with large-cap firms.

# 4 Estimating Counterfactual Closing Option Price

To estimate a counterfactual option price ($\hat{P}_t$) consistent with the stock closing auction price, we utilize the best-performing GBM model described earlier. This is achieved by setting other variables at their respective values at market close. The estimation equation is given by:

$$\widehat{p}_{auc} = f_{\text{GBM}}(S_{\text{auc}}, X; \theta). \tag{3}$$

Where $S_{\text{auc},t}$ represents the underlying stock price from the closing call auction. To examine the deviations in option prices at market close, we define the deviation as the percentage change in logarithmic form, following the approach of Bogousslavsky and Muravyev (2022):

$$\text{Deviation} = |log(p/\widehat{p}_{auc})|$$

Here, $p$ denotes the current option close price, the last transaction price, or the midquote at 4 PM.

## 4.1 Price Deviations from ML Benchmark

The distribution of estimated deviations by underlying firm size is presented in Table 8.

[Insert Table 7 Here]

In Table 7 Panel A, we examine the deviations of options' mid-quotes at 4 PM from our machine learning benchmark price. On average, the mid-quotes deviate by 35% from our benchmark, with a median deviation of 8.4%. Notably, this deviation is particularly pronounced for the first quintile (Small) size of S&P 500 firms, where the average deviation is 35%, and the 25th and 75th percentiles of deviations are 3.6% and 33%, respectively. In comparison, for the fifth quintile (Large) size of S&P 500 firms, the average deviation is 28%, and the 25th and 75th percentiles of deviations are 1.75% and 21%, respectively. Furthermore, it is worth noting that deviations for larger firms exhibit larger variance.

We investigate the deviations compared to the last traded price in Panel B. The deviations are even more pronounced in this case, with the average deviation increasing to 47% and a median deviation of 22.5%. A similar pattern is observed when considering the deviation of the last transaction price, where options associated with large firms, which have more frequently traded options, exhibit less deviation than smaller firms. Specifically, the average deviation for small firms is 48% while for large firms, it is 36%, and the respective medians are 27% and 13.8%. Furthermore, the interquartile range reveals that the options of large firms display less variation, with a range of 31 compared to 50 for small firms.

Note that the number of closing transactions is lower due to never-traded equity options. While our quote data encompasses all series of equity options regardless of trades, transactions may not have occurred or are unavailable for a fraction of the option series. Last transactions are more likely to be available for options associated with larger firms, which aligns with our findings above.

[Insert Table 8 Here]

Table 8 provides detailed insights into the estimated deviations based on the moneyness of options. The analysis reveals distinct patterns in the percentage deviations observed across different moneyness categories. Out-of-the-money (OTM) options exhibit substantially larger deviations compared to at-the-money (ATM) and in-the-money (ITM) options. The median deviation for OTM options is 43.5%, indicating a notable deviation from the benchmark price. Furthermore, the average deviation for OTM options is even more pronounced at 75.5%, underscoring the significance of the deviations observed in this category. In contrast, ATM options demonstrate a relatively lower median deviation of 11.4%, with an average deviation of 22.8%. Similarly, ITM options exhibit a median deviation of 2.59% and an average deviation of 3.88%.

The deviations observed for OTM options can be attributed to their lower price levels. As options move further out of the money, their prices decrease, leading to larger percentage deviations relative to the benchmark price. The substantial variations in deviations within the OTM category, as indicated by the interquartile range of 241.8, suggest significant dispersion in the deviations observed for individual OTM options. In contrast, ATM and ITM options exhibit relatively lower deviations, with narrower interquartile ranges of 20.9 and 3.98, respectively.

Shifting the focus to the deviations based on the last transaction price, an intriguing observation emerges. Compared to the deviations calculated using the 4 pm quote, expected to reflect more timely information at the close, the deviations based on the last transaction price are significantly greater for ATM and ITM options. Specifically, the average deviation for ATM options increases to 31.15%, representing a notable 36% increment from the corresponding deviation based on the 4 pm quote (22.8%). Similarly, the average deviation for ITM options surges to 22.79%, experiencing a substantial 587% increase from the 4 pm quote deviation (3.88%). This finding highlights a noteworthy vulnerability in the current reporting practices, particularly concerning ITM options, which possess greater value than their OTM counterparts. The substantial deviations observed for ITM options when relying on the last transaction price may introduce significant biases in the valuation of asset holdings and warrant careful consideration.

## 4.2 Trade Volume, Transaction Frequency and Deviations

We employ regression analysis to investigate the factors associated with deviations from the benchmark price. The regression model is specified as follows:

$$|log(p/\hat{p}_{auc})|_{i,j,t} = \beta_1 \log(\text{Volume})_{i,j,t} + \beta_2 \text{ZeroTrade}_{i,j,t} + X'\beta + \lambda_j + \gamma_t + e_{i,j,t}, \quad (4)$$

where $i$, $j$, and $t$ denote the option, underlying equity, and day, respectively. To account for time-invariant firm-specific characteristics and abnormal day effects, we include two-way firm and day fixed effects. Standard error clustering is applied to address potential heteroscedasticity. The control variables incorporated in the regression analysis encompass days to expiration, option moneyness, absolute delta, and gamma. Notably, the inclusion of absolute delta helps control for the moneyness of the option.

Price deviations in options can be attributed to various factors, including stale quotes and noise trades. The stale quote hypothesis suggests that if options are traded less frequently within a given day, the quotes provided by market makers may be less accurate, leading to deviations between the midquote value and the fair option price based on the stock closing auction price. On the other hand, deviations can also arise from noise trades, especially in the proximity of the market close.

To investigate these hypotheses, we employ proxies to capture the effects of stale quotes and noise trades. We use transaction volume and frequency traded during the day as proxies for stale quotes, as the lower trading activity may indicate a higher likelihood of stale quotes and subsequent price deviations. Additionally, we employ the accumulated trade volume and transaction count during the last 10 minutes of trading to capture the impact of noise trades occurring near the close. By incorporating these variables into our analysis, we aim to examine the influence of stale quotes and noise trades on the observed price deviations in options.

We analyze two separate regressions to examine the influence of trading volume and its relationship with price deviations in options. The first regression focuses on the transaction volume during the day as a proxy for stale quotes, while the second regression considers the volume specifically within the last minute of trading to capture the impact of noise trades near the close.

In the first regression, we include daily transaction volume as an independent variable to assess its association with price deviations. We aim to determine whether lower daily trading activity is linked to a higher likelihood of stale quotes and, subsequently, larger price deviations.

In the second regression, we introduce the volume of trades within the last minute of trading. This variable allows us to examine the influence of noise trades near the market close on price deviations. By distinguishing between the trading volume during the day and the volume at the last minute, we can assess their individual effects on price deviations and explore any potential correlations between these variables.

[Insert Table 9 Here]

The estimation results are presented in Table 9. To illustrate the extent of variable correlations, we construct regressions by progressively including additional variables. Remarkably, throughout the different specifications, we consistently find a positive relationship between options trading volume and deviation at the close, thus providing compelling evidence supporting the trade noise hypothesis.

Analyzing the deviations based on the 4 pm midquote, we find that put options exhibit lower deviations than call options. Furthermore, options with a longer time to maturity demonstrate a slight reduction in bias. Interestingly, we do not observe a statistically significant relationship between the realized volatility of the underlying asset and option pricing errors at the close.

Our analysis yields consistent estimates regarding the impact of moneyness on option deviations, as outlined earlier. Specifically, we observe that in-the-money (ITM) options exhibit lower deviations compared to at-the-money (ATM) and out-of-the-money (OTM) options. Moreover, when examining the 4 pm quotes, we identify a notable pattern of decreasing deviations as the size of the underlying firm increases. This pattern remains consistent across the top 15 most traded companies, where lower deviations are observed.

# 5 Conclusion

Our study thoroughly analyzes the challenges in determining closing prices for equity options and introduces a novel approach to address these issues. The study emphasizes the limitations of relying on proxy measures, such as last transaction prices and end-of-day mid-quotes, which can lead to stale and noisy information.

To mitigate these challenges, we develop a counterfactual benchmark that combines a machine learning algorithm with the closing auction price of the underlying stock, which is known for its reliability in the equity market; the proposed methodology enhances the accuracy and robustness of option pricing.

Our results reveal that our parsimonious machine-learning models surpass the accuracy of traditional models. This improvement can be attributed to capturing complex functional forms and feature interdependencies. Furthermore, our analysis demonstrates that traditional input variables retain significant information content in predicting option prices. Our feature importance and regression analysis on machine learning predictions reveal where the traditional model fails to account for. Volatility is deemed important in the BSM model, while time to maturity gains relative importance in the GBM model, indicating the varying impact of different features in predicting option prices.

In addition to its contributions to option pricing methodology, this study sheds light on implementing call auctions at the close in the options market. We found significant deviations in both the mid-quotes and last transaction prices of equity options when compared to our counterfactual machine learning benchmark based on the stock close auction price. These deviations are particularly pronounced for smaller-sized firms and out-of-the-money (OTM) options. Furthermore, the deviations of the last transaction price for in-the-money (ITM) options are even more substantial compared to the 4 PM mid-quote. Notably, the deviations in the last transaction price for in-the-money (ITM) options are even more substantial. These findings carry economic significance as they affect the accurate reporting of mutual fund holdings, given the reliance on historical prices, and the higher asset value associated with ITM options.

By highlighting the potential benefits of introducing call auctions for options, the research provides insights and analysis that can enhance the understanding of closing prices in the equity options market.

Overall, this research addresses the challenges associated with determining closing prices for equity options and presents a practical and effective solution using machine learning and the closing auction price of the underlying stock. The findings demonstrate the superiority of machine learning models over traditional models in terms of accuracy and provide interpretability analysis. By contributing to the literature on option pricing methodology and emphasizing the potential benefits of call auctions for options, this research offers valuable implications for both academia and industry practitioners.

# References

Admati, A. R. and P. Pfleiderer (1988). A theory of intraday patterns: Volume and price variability. *The review of financial studies 1*(1), 3–40.

Andreou, P. C., C. Han, and N. Li (2023, April). Stock options pricing via machine learning methods combined with firm characteristics.

Barbon, A., H. Beckmeyer, A. Buraschi, and M. Moerke (2021). Liquidity provision to leveraged etfs and equity options rebalancing flows: Evidence from end-of-day stock prices. *Swiss Finance Institute Research Paper* (22-40).

Barclay, M. J., T. Hendershott, and C. M. Jones (2008, March). Order consolidation, price efficiency, and extreme liquidity shocks. *Journal of Financial and Quantitative Analysis 43*(1), 93–121. Publisher: Cambridge University Press.

Barone-Adesi, G. and R. E. Whaley (1987). Efficient Analytic Approximation of American Option Values. *The Journal of Finance 42*(2), 301–320. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1987.tb02569.x.

Bergsma, K., A. Fodor, V. Singal, and J. Tayal (2020). Option trading after the opening bell and intraday stock return predictability. *Financial Management 49*(3), 769–804. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/fima.12284.

Black, F. and M. Scholes (1973, June). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy 81*(3), 637. Publisher: University of Chicago.

Bogousslavsky, V. (2021, July). The cross-section of intraday and overnight returns. *Journal of Financial Economics 141*(1), 172–194.

Bogousslavsky, V. and D. Muravyev (2022, May). Who trades at the close? implications for price discovery and liquidity. *Implications for Price Discovery and Liquidity (October 25, 2022).*

Broadie, M. and P. Glasserman (1997, June). Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control 21*(8), 1323–1352.

Budish, E., P. Cramton, and J. Shim (2015, November). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics 130*(4), 1547–1621.

Cao, J., J. Chen, and J. Hull (2020, September). A neural network approach to understanding implied volatility movements. *Quantitative Finance 20*(9), 1405–1413. Publisher: Routledge _eprint: https://doi.org/10.1080/14697688.2020.1750679.

Chan, K., Y. P. Chung, and H. Johnson (1995, September). The Intraday Behavior of Bid-Ask Spreads for NYSE Stocks and CBOE Options. *Journal of Financial and Quantitative Analysis 30*(3), 329–346. Publisher: Cambridge University Press.

Coval, J. D. and T. Shumway (2001). Expected option returns. *The journal of Finance 56*(3), 983–1009.

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

Ferguson, R. and A. D. Green (2018, October). Deeply Learning Derivatives.

Gu, S., B. Kelly, and D. Xiu (2020, May). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Heston, S. L. (1993, April). 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.

Hu, E. and D. Murphy (2021, October). Vestigial tails? floor brokers at the close in modern electronic markets. *Floor Brokers at the Close in Modern Electronic Markets (October 22, 2021)*.

Hull, J. and A. White (1987). The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance 42*(2), 281–300. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1987.tb02568.x.

Hutchinson, J. M., A. W. Lo, and T. Poggio (1994). A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks. *The Journal of Finance 49*(3), 851–889. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1994.tb00081.x.

Jegadeesh, N. and Y. Wu (2022, March). Closing auctions: Nasdaq versus nyse. *Journal of Financial Economics 143*(3), 1120–1139.

Longstaff, F. A. and E. S. Schwartz (2001, January). Valuing American Options by Simulation: A Simple Least-Squares Approach. *The Review of Financial Studies 14*(1), 113–147.

Malliaris, M. and L. Salchenberger (1993). A neural network model for estimating option prices. *Applied Intelligence 3*, 193–206.

Muravyev, D. (2016). Order flow and expected option returns. *The Journal of Finance 71*(2), 673–708.

Muravyev, D. and X. C. Ni (2020, April). Why do option returns change sign from day to night? *Journal of Financial Economics 136*(1), 219–238.

Muravyev, D. and N. D. Pearson (2020, November). Options trading costs are lower than you think. *The Review of Financial Studies 33*(11), 4973–5014.

Pagano, M. S. and R. A. Schwartz (2003, June). A closing call's impact on market quality at euronext paris. *Journal of Financial Economics 68*(3), 439–484.

Plante, S. (2017, November). Should corporate bond trading be centralized? *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Ruf, J. and W. Wang (2020, May). Neural networks for option pricing and hedging: a literature review. arXiv:1911.05620 [cs, q-fin, stat].

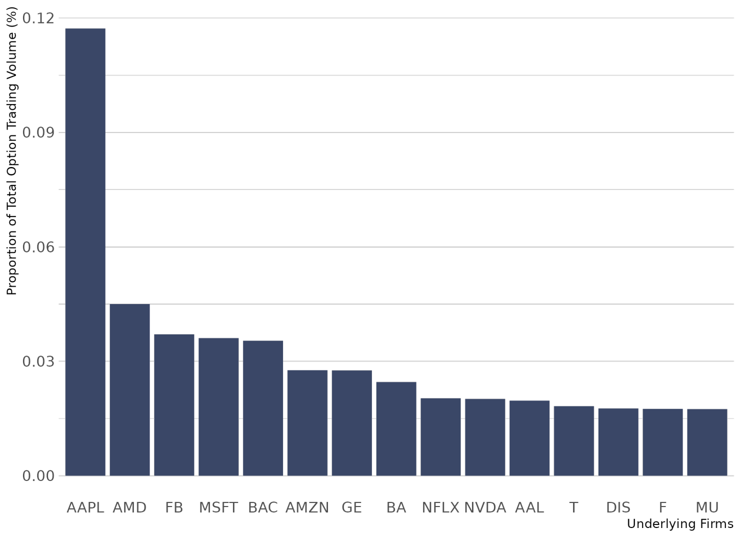
Scott, L. O. (1987, December). Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application. *Journal of Financial and Quantitative Analysis 22*(4), 419–438. Publisher: Cambridge University Press.

Wiggins, J. B. (1987, December). Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics 19*(2), 351–372.

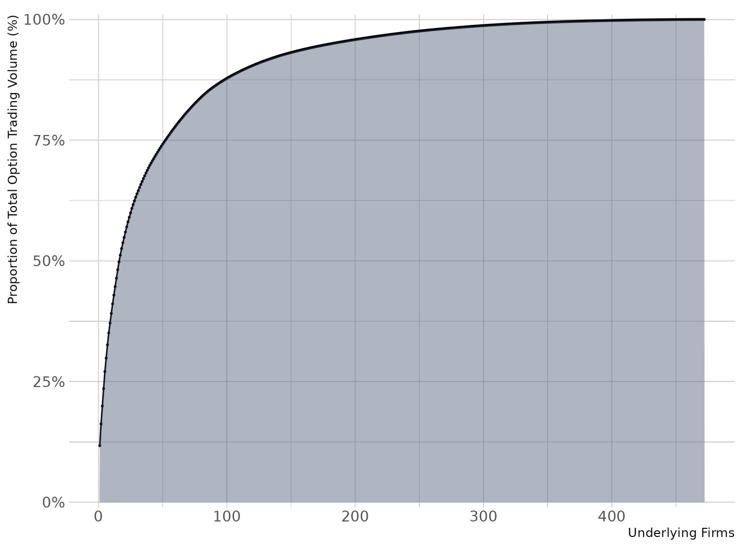
Wolpert, D. and W. Macready (1997, April). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation 1*(1), 67–82. Conference Name: IEEE Transactions on Evolutionary Computation.

Wood, R. A., T. H. McInish, and J. K. Ord (1985). An investigation of transactions data for nyse stocks. *The Journal of Finance 40*(3), 723–739.

# 6 Figures




(a) Top 15 Firms by Options Trading Volume




(b) Cumulative Volume S&P 500 Firms

Figure 1: Trading volume by Underlying Firms

*Note:* Figure (a) presents a visual representation of the top 15 most traded equity options in the United States during the sample period spanning from October 2019 to March 2021. Each column in the figure corresponds to the proportion of options trading volume for each firm. Figure (b) depicts the cumulative distribution of the proportional transaction volume encompassing all underlying firms.

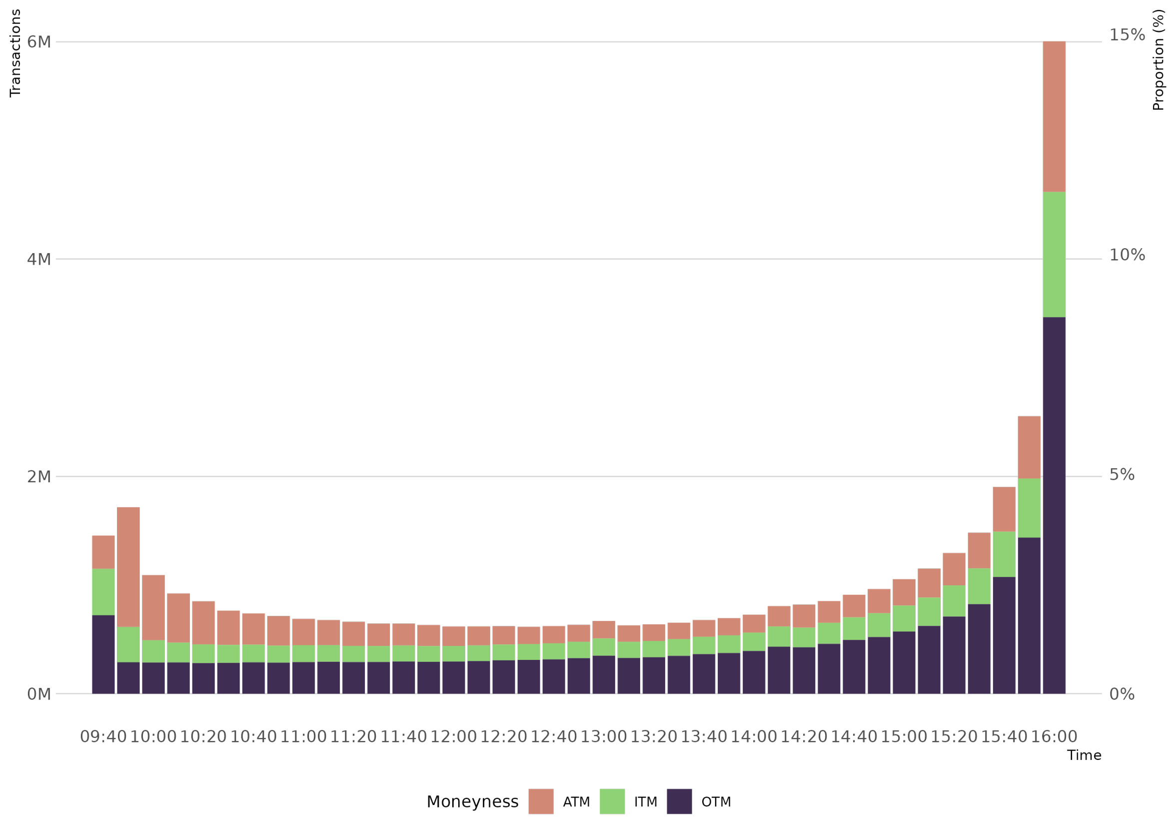
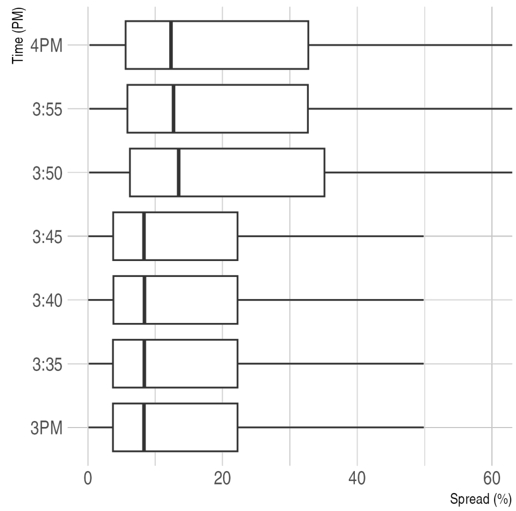
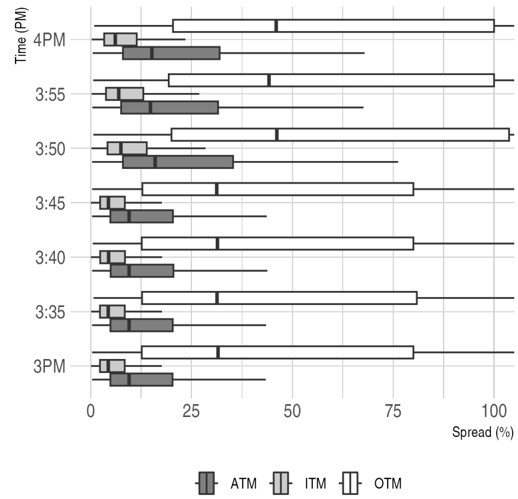



Figure 2: Options Transactions Distribution by Last Trading Time

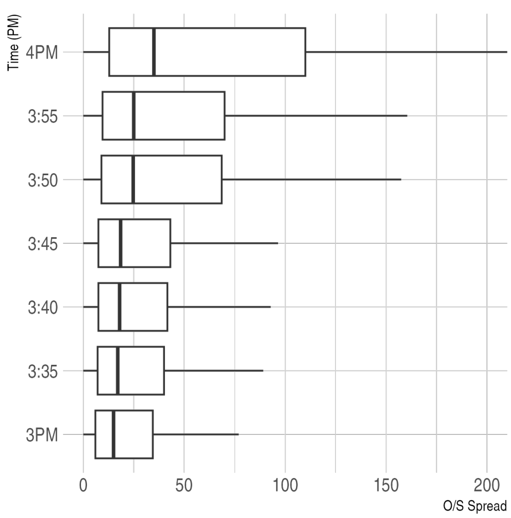
*Note:* The depicted figure demonstrates the stacked distribution of the final transaction time for each option series across trading days and moneyness. The horizontal axis represents transaction time categorized into intervals of 10 minutes, while the vertical axis denotes the summation of distinct options series. The classification of options' moneyness is established through a 10% threshold applied to the ratio between the stock price and strike price.



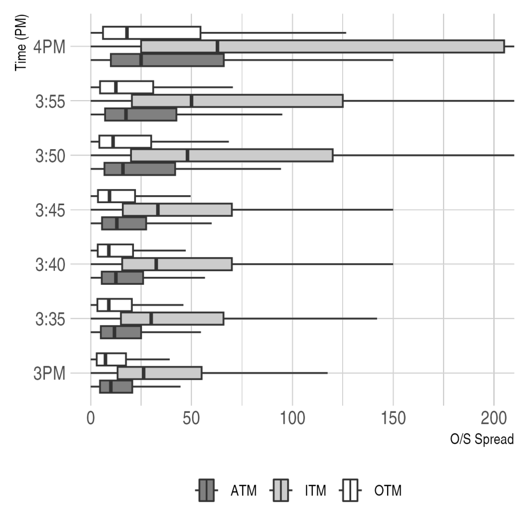

(a) Bid-ask spread




(b) Bid-ask spread, by moneyness




(c) Option to stock spread ratio




(d) Option to stock spread ratio, by moneyness

Figure 3: Intraday Quote Behavior near Close

*Note:* Above presents boxplots depicting the distribution of relative bid-ask spreads for options and the corresponding options-to-stocks bid-ask spread ratios across different intraday trading times. Options-to-stocks bid-ask ratio are calculated by dividing relative spread of options by that of underlying stocks. The boxplots represent the data's minimum, maximum, median, and interquartile range. The classification of options' moneyness is established through a 10% threshold applied to the ratio between the stock price and strike price. To optimize computational efficiency, a random sample of one million observations was drawn from the comprehensive dataset comprising all S&P 500 equity options in 2020. We exclude records with zero bid quotes due to the mechanical result of a 200% spread.

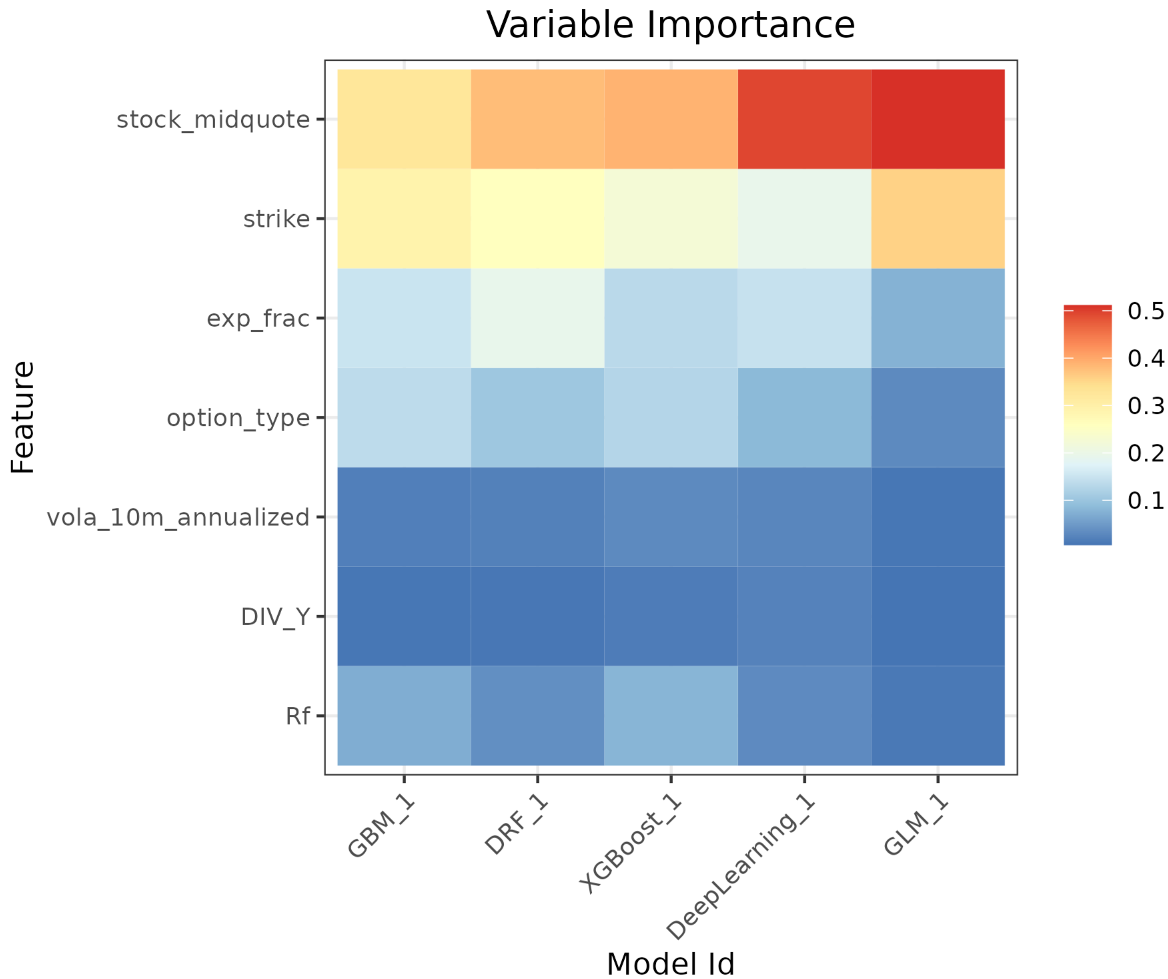
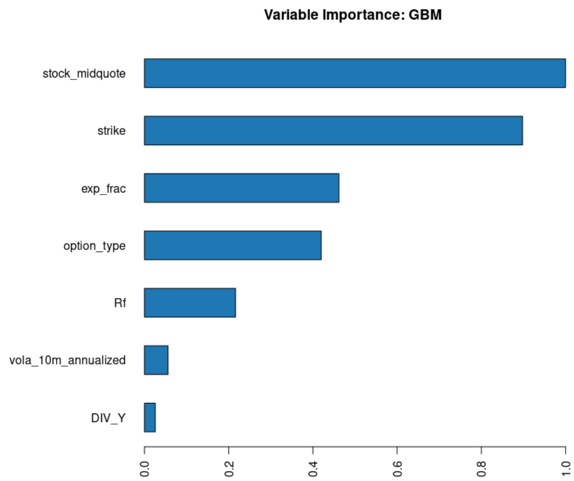
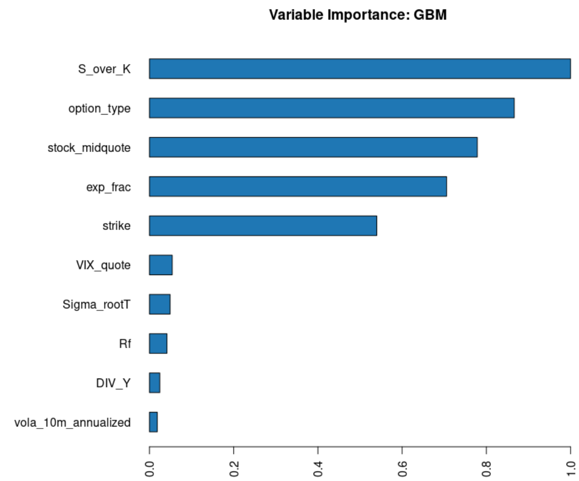



Figure 4: Feature Importance Heatmap with BSM Input Variables

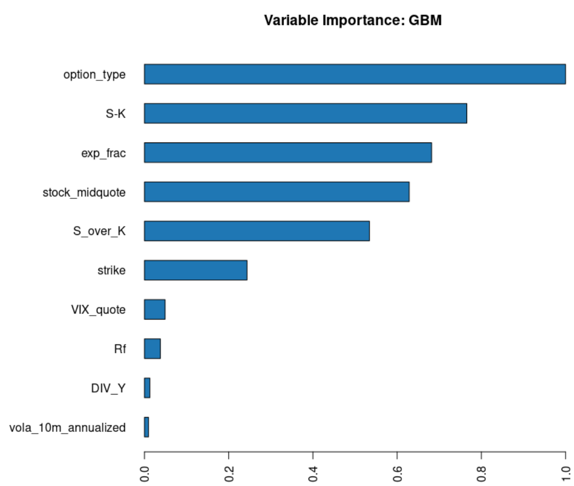
*Note:* The above figure provides a visual representation of the relative significance of input variables (features) in various machine learning algorithms. The vertical axis displays the input variables, while the horizontal axis represents each algorithm. Feature importance metrics quantify the relative influence or contribution of these input variables to the overall predictive performance of the models. Note that all algorithms receive identical input variables, which are also utilized in the computation of classical BSM and CRR models. The degree of importance is assessed within each individual algorithm.



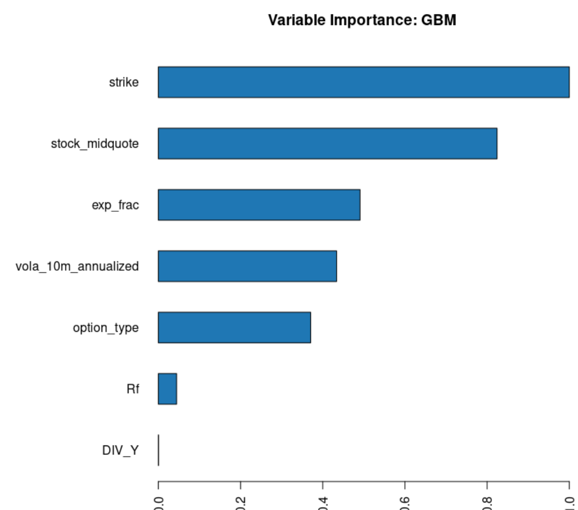

(a) GBM, with BSM input




(b) GBM, with weighted Volatility



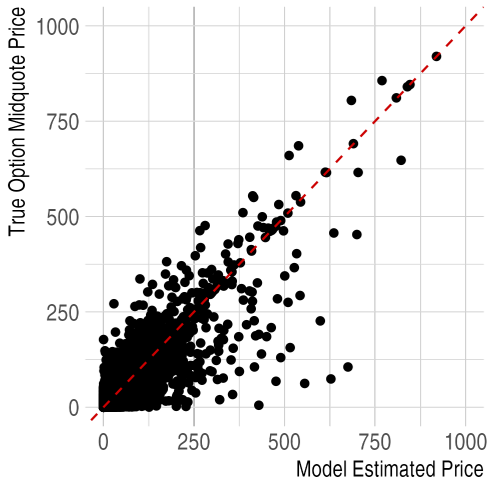

(c) GBM, with moneyness and VIX



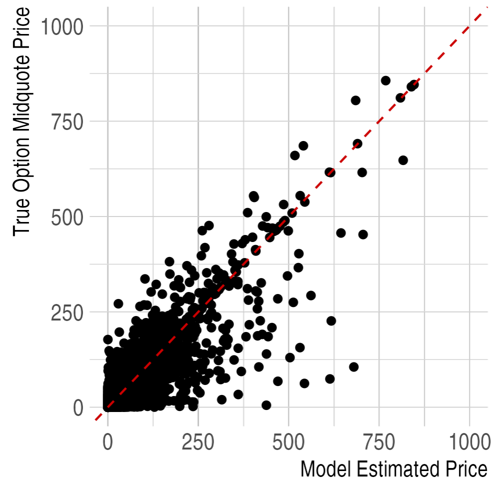

(d) GBM, Black-Scholes Counterfeit

Figure 5: Feature Importance Plots

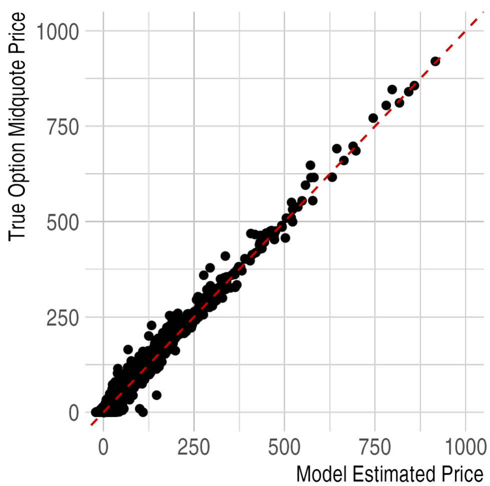
*Note:* The above figures depict the relative significance of input variables in each specifications for Boosted-Tree algorithm (GBM). The vertical axis displays the input variables, and the horizontal axis represents relative importance, scaling to 1 for the most important feature. Feature importance metrics quantify the contribution of these input variables to the overall predictive performance. Weighted volatility is calculated by multiplying annualized realized volatility ($\sigma$) and time to maturity (in years, $\tau$). Figure (d) display feature importance when the target variable is Black-Scholes estimated price.



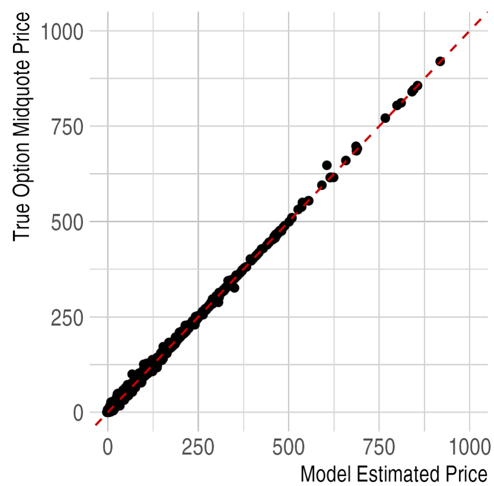

(a) BSM prediction




(b) CRR prediction




(c) GBM prediction




(d) GBM3 prediction

Figure 6: Accuracy Plots

*Note:* The plotted figures illustrate target (mid-quote price at the transaction) and the estimated option prices generated by each model. The data used in the above plot comprises transactions of S&P 500 equity options from October 2019 to March 2021. For efficiency, a random sample of 100,000 observations was drawn from the dataset for plotting. Any extreme values or zero predictions produced by the Black-Scholes-Merton (BSM) and Cox-Ross-Rubinstein (CRR) models, which signify divergence, were excluded from the plots. GBM3 model includes absolute $(S-K)$ and relative $(S/K)$ moneyness, VIX and standard option pricing variables.

# 7 Tables

Table 1: Descriptive Statistics

| Variable | N | Mean | SD | p1 | p25 | p50 | p75 | p99 |
|---|---|---|---|---|---|---|---|---|
| strike | 716,822,161 | 345.81 | 666.77 | 10.50 | 67.07 | 144.62 | 274.93 | 3,283.52 |
| trade_size | 716,822,161 | 6.32 | 81.85 | 1.00 | 1.00 | 1.78 | 4.62 | 61.90 |
| trade_price | 716,822,161 | 6.65 | 25.82 | 0.03 | 0.68 | 2.35 | 5.00 | 79.66 |
| option_midquote | 716,822,161 | 6.65 | 25.27 | 0.02 | 0.72 | 1.97 | 5.38 | 86.02 |
| stock_midquote | 716,822,161 | 343.23 | 657.34 | 8.73 | 68.08 | 143.23 | 274.11 | 3,209.23 |
| days_to_expire | 716,822,161 | 41.28 | 93.69 | 0.00 | 3.78 | 11.34 | 32.80 | 510.75 |
| DIV_Y | 716,730,281 | 0.02 | 0.02 | 0.00 | 0.00 | 0.01 | 0.03 | 0.07 |
| Rf | 710,020,847 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.02 | 0.02 |
| VIX_quote | 666,952,310 | 24.91 | 9.45 | 12.17 | 17.67 | 23.35 | 29.38 | 53.72 |
| ba_pspread | 716,784,655 | 0.11 | 0.26 | 0.00 | 0.02 | 0.04 | 0.09 | 1.74 |
| vola_10m_annualized | 707,469,043 | 0.40 | 0.46 | 0.09 | 0.23 | 0.31 | 0.49 | 1.81 |
| bsm_impl_vol | 702,950,708 | 0.49 | 0.33 | 0.16 | 0.30 | 0.42 | 0.58 | 1.89 |
| bsm_delta | 700,154,923 | 0.13 | 0.41 | -0.92 | -0.06 | 0.13 | 0.42 | 1.00 |
| bsm_gamma | 700,154,923 | 0.07 | 0.22 | 0.00 | 0.01 | 0.03 | 0.07 | 0.68 |
| bsm_vega | 700,154,923 | 15.61 | 45.72 | 0.00 | 1.60 | 5.49 | 14.33 | 188.71 |
| bsm_theta | 700,154,923 | -155.42 | 740.01 | -2,359.26 | -83.23 | -24.99 | -9.00 | 0.64 |

*Note:*
Each column represents nonmissing observations (N), mean, standard deviation (S.D.), and percentiles. The sample is constructed from OPRA options transactions data from October 2019 to March 2021 and consists of all listed S&P 500 equity options during the sample period. VIX quotes are only paired when the most recent quote data is accessible within a 500-second timeframe before the option transaction. The underlying realized volatility is computed by utilizing the 1-second return and subsequently annualizing it. Implied volatility and base Greeks are derived from the Black-Scholes model.

Table 2: Machine Learning Algorithm Estimation Accuracy Rankings

| Variable | GBM | RF | NN | XGB | EN | BSM | CRR | BAW | CN |
|---|---|---|---|---|---|---|---|---|---|
| **Mean Absolute Error** | | | | | | | | | |
| MAE_on_training | 1.268 | 2.363 | 2.557 | 2.464 | 6.178 | | | | |
| MAE_on_validation | 1.273 | 2.339 | 2.655 | 2.471 | 6.178 | | | | |
| MAE_on_test | 1.269 | 2.335 | 2.653 | 2.465 | 6.177 | 2.622 | 2.789 | 2.759 | 2.784 |
| **Root Mean Squared Error** | | | | | | | | | |
| RMSE_on_training | 2.891 | 5.988 | 6.453 | 10.339 | 22.729 | | | | |
| RMSE_on_validation | 2.940 | 5.795 | 7.619 | 10.404 | 22.747 | | | | |
| RMSE_on_test | 2.904 | 5.749 | 7.592 | 10.339 | 22.726 | 10.871 | 12.718 | 12.153 | 12.732 |
| **Out of Sample R-squared** | | | | | | | | | |
| r2_on_training | 0.987 | 0.944 | 0.901 | 0.833 | 0.193 | | | | |
| r2_on_validation | 0.987 | 0.948 | 0.909 | 0.831 | 0.193 | | | | |
| r2_on_test | 0.987 | 0.948 | 0.910 | 0.833 | 0.193 | 0.817 | 0.709 | 0.772 | 0.708 |
| **Mean Percentage Absolute Error (Test sample)** | | | | | | | | | |
| MPAE_on_test | 56.030 | 80.713 | 79.110 | 59.387 | 103.939 | 196.830 | 95.496 | 183.854 | 127.738 |

*Note:*

The training, validation, and test datasets encompass all equity options listed in the S&P 500 index from October 2019 to March 2021. The examined algorithms include Gradient Boosted Machine (GBM), Random Forest (RF), Neural Network (feed-forward, NN), Extreme Gradient Boosting (XGB), Elastic Net (EN), Black-Scholes-Merton (BSM), Cox-Ross-Rubinstein (CRR), Barone-Adesi and Whaley (BAW), and Crank-Nicolson (CN). The mean percentage absolute error (MPAE) is computed by

$$\frac{1}{N}\sum_{i=1}^{N} 100 * |log(\frac{\hat{y}_i}{y_i})|$$

To mitigate extreme values resulting from divergence, the CRR and CN estimates are winsorized at the 0.01% level.

Table 3: Machine Learning Algorithm Estimation Accuracy Rankings, by Moneyness

| Variable | GBM | RF | NN | XGB | EN | BSM | CRR | BAW | CN |
|---|---|---|---|---|---|---|---|---|---|
| **OTM Options** | | | | | | | | | |
| MAE_on_OTM | 1.025 | 1.909 | 2.037 | 1.834 | 4.836 | 2.061 | 2.064 | 2.060 | 2.058 |
| RMSE_on_OTM | 2.064 | 3.319 | 6.025 | 5.923 | 10.890 | 8.505 | 8.257 | 8.507 | 8.276 |
| r2_on_OTM | 0.949 | 0.867 | 0.562 | 0.576 | -0.432 | 0.259 | 0.270 | 0.260 | 0.258 |
| MPAE_on_OTM | 78.448 | 114.122 | 91.816 | 77.871 | 130.348 | 327.639 | 148.204 | 300.991 | 202.924 |
| **ATM Options** | | | | | | | | | |
| MAE_on_ATM | 1.187 | 2.041 | 3.102 | 2.549 | 5.914 | 3.625 | 4.041 | 4.037 | 4.036 |
| RMSE_on_ATM | 2.800 | 5.727 | 8.047 | 8.860 | 20.426 | 12.865 | 15.843 | 15.866 | 15.817 |
| r2_on_ATM | 0.987 | 0.948 | 0.897 | 0.875 | 0.333 | 0.702 | 0.547 | 0.549 | 0.547 |
| MPAE_on_ATM | 25.059 | 34.887 | 61.757 | 33.553 | 65.718 | 43.103 | 44.148 | 44.240 | 44.327 |
| **ITM Options** | | | | | | | | | |
| MAE_on_ITM | 2.859 | 5.512 | 4.403 | 5.543 | 14.040 | 2.291 | 2.500 | 2.251 | 2.502 |
| RMSE_on_ITM | 5.918 | 12.356 | 12.082 | 24.479 | 54.724 | 14.247 | 18.927 | 14.223 | 19.043 |
| r2_on_ITM | 0.990 | 0.956 | 0.958 | 0.827 | 0.134 | 0.942 | 0.868 | 0.943 | 0.866 |
| MPAE_on_ITM | 39.822 | 55.495 | 69.277 | 46.984 | 90.438 | 15.541 | 15.346 | 15.345 | 15.334 |

*Note:*

All evaluation metrics are based on the test sample. Options with $|\Delta| < 0.375$ are classified as out-of-the-money (OTM), those with $|\Delta| >= 0.625$ are classified as in-the-money (ITM), and the rest are considered at-the-money (ATM). The algorithms encompass Gradient Boosted Machine (GBM), Random Forest (RF), Neural Network (feed-forward, NN), Extreme Gradient Boosting (XGB), Elastic Net (EN), Black-Scholes-Merton (BSM), Cox-Ross-Rubinstein (CRR), Barone-Adesi and Whaley (BAW), and Crank-Nicolson (CN). The mean percentage absolute error (MPAE) is calculated using the following formula:

$$\frac{1}{N}\sum_{i=1}^{N} 100 * |log(\frac{\hat{y}_i}{y_i})|$$

To mitigate extreme values resulting from divergence, the CRR and CN estimates are winsorized at the 0.01% level.

Table 4: GBM Estimation Accuracy Improvements

Panel A: Performance Improvements

| Variable | GBM | GBM2 | GBM3 |
|---|---|---|---|
| **Mean Absolute Error** | | | |
| MAE_on_training | 1.268 | 0.580 | 0.394 |
| MAE_on_validation | 1.273 | 0.583 | 0.398 |
| MAE_on_test | 1.276 | 0.584 | 0.395 |
| **Root Mean Squared Error** | | | |
| RMSE_on_training | 2.891 | 1.283 | 0.837 |
| RMSE_on_validation | 2.940 | 1.309 | 0.869 |
| RMSE_on_test | 2.960 | 1.309 | 0.850 |
| **Out of Sample R-squared** | | | |
| r2_on_training | 0.987 | 0.997 | 0.999 |
| r2_on_validation | 0.987 | 0.997 | 0.999 |
| r2_on_test | 0.987 | 0.997 | 0.999 |
| **Mean Percentage Absolute Error** | | | |
| MPAE_on_test | 56.030 | 28.101 | 21.243 |

Panel B: Performance by Moneyness

| Variable | GBM | GBM2 | GBM3 |
|---|---|---|---|
| **OTM Options** | | | |
| MAE_on_OTM | 1.025 | 0.463 | 0.317 |
| RMSE_on_OTM | 2.064 | 1.002 | 0.660 |
| r2_on_OTM | 0.949 | 0.988 | 0.995 |
| MPAE_on_OTM | 78.448 | 38.789 | 29.507 |
| **ATM Options** | | | |
| MAE_on_ATM | 1.187 | 0.716 | 0.529 |
| RMSE_on_ATM | 2.800 | 1.496 | 1.092 |
| r2_on_ATM | 0.987 | 0.996 | 0.998 |
| MPAE_on_ATM | 25.059 | 16.599 | 12.881 |
| **ITM Options** | | | |
| MAE_on_ITM | 2.859 | 0.783 | 0.370 |
| RMSE_on_ITM | 5.918 | 1.924 | 0.832 |
| r2_on_ITM | 0.990 | 0.999 | 1.000 |
| MPAE_on_ITM | 39.822 | 9.862 | 5.426 |

*Note:*
The results are based on a testing sample of S&P 500 equity options from October 2019 to March 2021. In our first modification (GBM2), we include the S&P 500 VIX index quote and time-weighted volatility ($\sigma\sqrt{\tau}$) following the Black-Scholes-Merton (BSM) model. The underlying volatility ($\sigma$) is weighted by the square root of the time-to-maturity ($\sqrt{\tau}$), as illustrated in the BSM model equation. In our second modification (GBM3), we introduce the absolute difference ($S-K$) and relative moneyness ($S/K$) of the option instead of the weighted time-to-maturity.

Table 6: Determinants of Prediction Errors

| Dependent Variables: | bsm_abs_deviation*100 | | gbm_abs_deviation*100 | |
|---|---|---|---|---|
| Model: | (1) | (2) | (3) | (4) |
| *Variables* | | | | |
| moneyness_aITM | -8.33 | 54.87*** | -14.72*** | -9.39*** |
| | (-0.70) | (2.93) | (-13.38) | (-13.30) |
| moneyness_aOTM | 354.49*** | 441.99*** | 11.99*** | 19.47*** |
| | (9.47) | (10.24) | (26.85) | (29.80) |
| option_typeP | -44.26*** | -50.28*** | -0.65** | -1.19*** |
| | (-2.99) | (-3.40) | (-2.44) | (-4.32) |
| log(1+trade_size) | 11.37*** | 10.90*** | 1.09*** | 1.05*** |
| | (3.48) | (3.59) | (14.40) | (17.84) |
| trade_price | -2.38*** | -1.65*** | -0.10* | -0.04 |
| | (-4.93) | (-8.40) | (-1.89) | (-1.48) |
| stock_midquote | 0.01 | -0.01 | 0.00 | 0.00*** |
| | (0.15) | (-0.23) | (-1.15) | (-3.76) |
| I(100 × vola_10m_annualized) | -0.62** | -0.66** | 0.00 | 0.00 |
| | (-2.29) | (-2.31) | (1.62) | (-1.11) |
| I(Rf × 100) | -158.83*** | -77.27*** | -7.17*** | -0.64 |
| | (-6.65) | (-9.48) | (-4.32) | (-0.65) |
| I(DIV_Y × 100) | 0.83 | 0.78 | -0.22** | -0.24** |
| | (0.85) | (0.89) | (-2.20) | (-2.27) |
| exp_frac | | -859.02*** | | -69.74*** |
| | | (-12.48) | | (-37.03) |
| exp_frac square | | 446.82*** | | 35.25*** |
| | | (12.19) | | (36.59) |
| firmsizeMedium | | -40.94*** | | 1.08 |
| | | (-4.35) | | (0.83) |
| firmsizeMid-Large | | -30.30*** | | 0.22 |
| | | (-4.01) | | (0.63) |
| firmsizeMid-Small | | -53.41*** | | -0.09 |
| | | (-3.95) | | (-0.07) |
| firmsizeSmall | | -56.29*** | | -1.54 |
| | | (-3.10) | | (-1.09) |
| *Fixed-effects* | | | | |
| root | Yes | Yes | Yes | Yes |
| quote_date | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | |
| Observations | 48,530,785 | 48,528,073 | 48,134,720 | 48,132,018 |
| Adjusted $R^2$ | 0.03253 | 0.03902 | 0.08808 | 0.13297 |

*Clustered (root) co-variance matrix, t-stats in parentheses*
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The regression analysis is conducted using a randomly selected sample of 50 million observations from a larger dataset consisting of 716 million transactions. The primary dependent variable is the percentage log deviation of the estimate from the midpoint of the option quote. Both the risk-free rate and dividend yield are expressed as percentages. The variable "exp_frac" represents the expiration period measured in years, with a value of 0.5 indicating half a year. The classification of options based on moneyness is determined using the absolute Black-Scholes delta. Options with $|\Delta| < 0.375$ are classified as out-of-the-money (OTM), those with $|\Delta| >= 0.625$ are classified as in-the-money (ITM), and the rest are considered at-the-money (ATM). The reference levels for the analysis are call option (option type) and ATM (moneyness).

Table 7: Price Deviations from Machine-Learning Benchmark by the Size of Firm

Panel A: Deviations from ML Benchmark, 4 PM Mid-quote

| Variable | All | Small | Mid-Small | Mid | Mid-Large | Large |
|---|---|---|---|---|---|---|
| **Quantiles** | | | | | | |
| 1% | 0.09 | 0.13 | 0.11 | 0.10 | 0.08 | 0.06 |
| 5% | 0.46 | 0.64 | 0.54 | 0.48 | 0.42 | 0.31 |
| 25% | 2.63 | 3.60 | 3.05 | 2.72 | 2.30 | 1.75 |
| 50% | 8.46 | 10.92 | 9.50 | 8.70 | 6.95 | 5.33 |
| 75% | 33.39 | 38.01 | 34.92 | 32.88 | 26.55 | 21.28 |
| 95% | 171.28 | 158.14 | 149.34 | 152.71 | 145.73 | 141.55 |
| 99% | 357.52 | 307.98 | 296.69 | 300.91 | 316.49 | 366.68 |
| **Statistics** | | | | | | |
| Mean | 35.53 | 35.47 | 33.10 | 32.70 | 30.07 | 28.72 |
| SD | 69.59 | 62.03 | 59.57 | 60.72 | 61.90 | 68.26 |
| Nobs | 70,814,372 | 13,801,742 | 13,949,480 | 13,995,909 | 13,820,873 | 13,807,125 |

Panel B: Deviations from ML Benchmark, Last Transaction Price

| Variable | All | Small | Mid-Small | Mid | Mid-Large | Large |
|---|---|---|---|---|---|---|
| **Quantiles** | | | | | | |
| 1% | 0.28 | 0.38 | 0.38 | 0.36 | 0.27 | 0.18 |
| 5% | 1.42 | 1.92 | 1.92 | 1.84 | 1.34 | 0.89 |
| 25% | 8.26 | 10.67 | 10.67 | 10.23 | 7.57 | 5.07 |
| 50% | 22.49 | 27.15 | 27.14 | 26.16 | 20.15 | 13.88 |
| 75% | 55.46 | 60.80 | 61.16 | 59.86 | 49.55 | 36.32 |
| 95% | 180.28 | 170.25 | 170.26 | 171.43 | 162.84 | 148.87 |
| 99% | 349.77 | 309.52 | 308.52 | 310.08 | 326.44 | 372.29 |
| **Statistics** | | | | | | |
| Mean | 47.12 | 48.56 | 48.66 | 48.09 | 42.65 | 36.47 |
| SD | 69.30 | 62.51 | 62.57 | 63.04 | 63.94 | 68.34 |
| Nobs | 45,288,774 | 8,272,550 | 8,101,959 | 8,280,355 | 8,864,817 | 10,830,411 |

*Note:*
The table presented above provides a summary of deviation estimates categorized by the size of the underlying firm. The relative size of firms within the S&P 500 index is classified into quintiles based on their daily market capitalization. Panel A displays the deviation statistics for the 4 PM mid-quote value, using end-of-day quotes sourced from CBOE options quote data covering the year 2020. Panel B shows the deviations for the most recent transaction price matched at the close of each trading day. The deviations are presented in logarithmic percentage values as $100 \times \log\left(\frac{y}{y_{ML}}\right)$.

Table 8: Price Deviations from Machine-Learning Benchmark by Moneyness of Options

Panel A: Deviations from ML Benchmark, 4 PM Mid-quote

| Variable | All | OTM | ATM | ITM |
|---|---|---|---|---|
| **Quantiles** | | | | |
| 1% | 0.09 | 0.6 | 0.2 | 0.04 |
| 5% | 0.46 | 3.0 | 0.8 | 0.22 |
| 25% | 2.63 | 17.0 | 4.6 | 1.14 |
| 50% | 8.46 | 43.5 | 11.4 | 2.59 |
| 75% | 33.39 | 98.8 | 25.5 | 5.12 |
| 95% | 171.28 | 258.8 | 84.3 | 11.89 |
| 99% | 357.52 | 439.6 | 185.0 | 19.98 |
| **Statistics** | | | | |
| Mean | 35.53 | 75.5 | 22.8 | 3.88 |
| SD | 69.59 | 91.0 | 36.2 | 4.34 |
| Nobs | 70,814,372 | 21,776,855 | 20,798,710 | 26,872,505 |

Panel B: Deviations from ML Benchmark, Last Transaction Price

| Variable | All | OTM | ATM | ITM |
|---|---|---|---|---|
| **Quantiles** | | | | |
| 1% | 0.28 | 0.63 | 0.28 | 0.16 |
| 5% | 1.42 | 3.18 | 1.41 | 0.80 |
| 25% | 8.26 | 17.49 | 7.62 | 4.44 |
| 50% | 22.49 | 43.59 | 18.48 | 11.57 |
| 75% | 55.46 | 96.10 | 39.10 | 27.79 |
| 95% | 180.28 | 251.98 | 103.80 | 84.05 |
| 99% | 349.77 | 428.34 | 194.30 | 152.38 |
| **Statistics** | | | | |
| Mean | 47.12 | 74.02 | 31.15 | 22.79 |
| SD | 69.30 | 88.06 | 39.69 | 31.57 |
| Nobs | 45,288,774 | 16,119,441 | 15,785,127 | 12,487,067 |

*Note:*
The table presented above provides a summary of deviation estimates categorized by moneyness of options. The option's moneyness is defined based on +/- 10% of stock-to-strike ratio (S/K). Panel A (B) displays the deviation statistics for the 4 PM mid-quote (most recent traded price). The deviations are presented in logarithmic percentage values as $100 \times \log\left(\frac{y}{y_{ML}}\right)$.

Table 9: Determinants of Benchmark Deviations

| Dependent Variables: | log_dev_4pm_pct | | log_dev_last_pct | |
|---|---|---|---|---|
| Model: | (1) | (2) | (3) | (4) |
| *Variables* | | | | |
| option_typeP | -3.72*** | -4.46*** | -0.99** | -3.61*** |
| | (-8.53) | (-11.63) | (-1.97) | (-8.20) |
| log_under_dev_pct | -0.01 | 0.00 | -0.03 | -0.03 |
| | (-0.35) | (-0.14) | (-1.13) | (-1.21) |
| vola_10m_annualized | 0.00 | 0.00 | 0.00 | 0.00 |
| | (-0.53) | (-0.49) | (-1.18) | (-1.04) |
| moneyness_aITM | -14.84*** | -18.91*** | 0.37 | -14.62*** |
| | (-29.94) | (-48.05) | (0.66) | (-34.14) |
| moneyness_aOTM | 59.81*** | 57.80*** | 52.19*** | 46.47*** |
| | (114.47) | (117.89) | (88.84) | (85.48) |
| days_expire_fct<30 | -26.13*** | -26.78*** | -28.84*** | -29.92*** |
| | (-20.10) | (-19.73) | (-20.73) | (-20.20) |
| days_expire_fct<90 | -37.05*** | -38.14*** | -43.83*** | -45.51*** |
| | (-25.45) | (-24.45) | (-30.26) | (-27.91) |
| days_expire_fct90+ | -50.04*** | -51.16*** | -57.43*** | -62.34*** |
| | (-31.79) | (-29.65) | (-38.48) | (-34.09) |
| no_trade1 | | 8.29*** | | 15.16*** |
| | | (16.15) | | (31.32) |
| log(1+trade_volume) | | -0.92*** | | -1.36*** |
| | | (-3.77) | | (-5.66) |
| days_since_last_trade | | | | 0.23*** |
| | | | | (34.21) |
| *Fixed-effects* | | | | |
| root | Yes | Yes | Yes | Yes |
| quote_date | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | |
| Observations | 69,160,875 | 69,160,875 | 44,231,044 | 44,231,044 |
| Adjusted $R^2$ | 0.29774 | 0.30179 | 0.21171 | 0.24805 |

*Clustered (root) co-variance matrix, t-stats in parentheses*
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

The panel regression is specified with underlying stock and day fixed effects paired with standard error clustering. The sample is based on S&P 500 quote data for the year 2020. The option's moneyness is defined based on +/- 10% of intrinsic value (S/K). The main dependent variables are the percentage deviation of the estimate from the ML benchmark option price. The base is call option and ATM.

# 8 Appendix

This page was intentionally left blank.

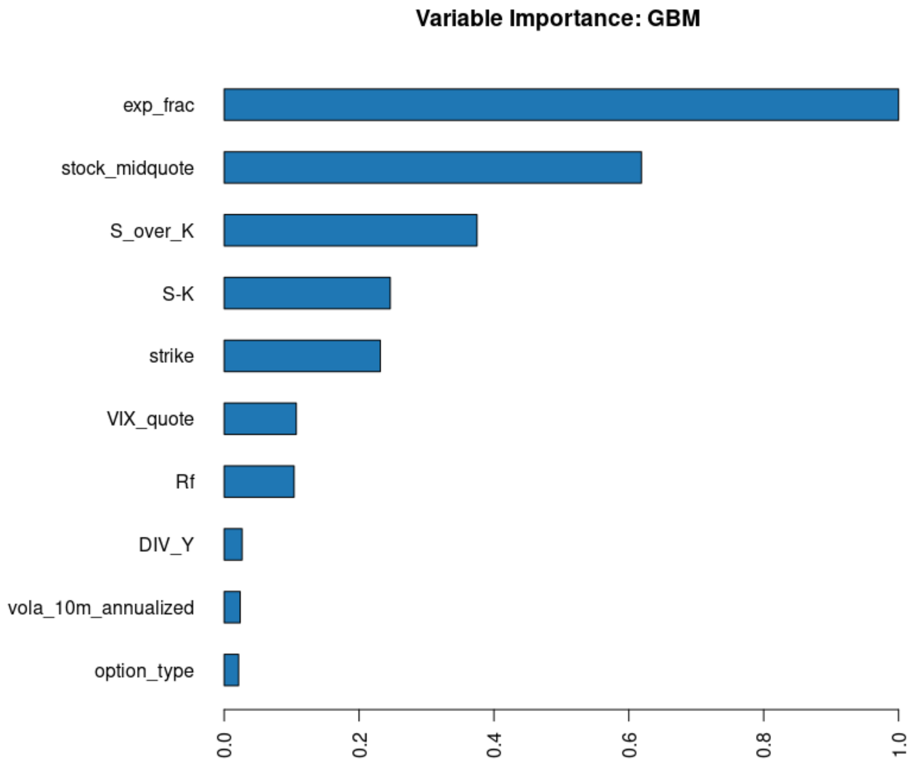



Figure 7: OTM Sample ML Variable importance

*Note:* The above figure depict the relative significance of input variables in extended GBM specification, including absolute (S-K) and relative (S/K) intrinsic values and VIX quote. We limit our train, valid and test sample to be OTM options. The vertical axis displays the input variables, and the horizontal axis represents relative importance, scaling to 1 for the most important feature. Feature importance metrics quantify the contribution of these input variables to the overall predictive performance.

Table A1: Estimation Accuracy Results by Option Price

| Variable | GBM | GBM2 | GBM3 | BSM | CRR | BAW | CN |
|---|---|---|---|---|---|---|---|
| **Options price below $1** | | | | | | | |
| MAE_on_below1 | 0.665 | 0.225 | 0.147 | 0.416 | 0.423 | 0.420 | 0.420 |
| RMSE_on_below1 | 1.251 | 0.472 | 0.237 | 1.854 | 1.762 | 1.856 | 1.782 |
| r2_on_below1 | -17.723 | -1.672 | 0.330 | -40.101 | -36.137 | -40.205 | -37.007 |
| MPAE_onbelow1 | 106.557 | 50.319 | 38.370 | 391.093 | 142.005 | 353.315 | 214.104 |
| **Options price below $3** | | | | | | | |
| MAE_on_below3 | 0.727 | 0.369 | 0.276 | 1.182 | 1.204 | 1.197 | 1.197 |
| RMSE_on_below3 | 1.231 | 0.555 | 0.406 | 3.392 | 3.232 | 3.400 | 3.289 |
| r2_on_below3 | -3.704 | 0.044 | 0.488 | -34.718 | -31.435 | -34.885 | -32.591 |
| MPAE_onbelow3 | 34.325 | 20.696 | 15.769 | 112.670 | 89.326 | 112.852 | 102.493 |
| **Options price below $10** | | | | | | | |
| MAE_on_below10 | 1.286 | 0.692 | 0.507 | 2.692 | 2.753 | 2.745 | 2.746 |
| RMSE_on_below10 | 1.942 | 1.006 | 0.759 | 5.989 | 5.903 | 6.023 | 5.995 |
| r2_on_below10 | -0.041 | 0.721 | 0.841 | -8.892 | -8.609 | -9.006 | -8.911 |
| MPAE_onbelow10 | 25.387 | 13.897 | 10.237 | 76.368 | 65.926 | 76.849 | 71.953 |
| **Options price above $10** | | | | | | | |
| MAE_on_above10 | 4.104 | 1.841 | 1.148 | 11.778 | 12.890 | 12.697 | 12.889 |
| RMSE_on_above10 | 7.127 | 3.129 | 2.003 | 28.544 | 34.000 | 32.277 | 33.999 |
| r2_on_above10 | 0.987 | 0.997 | 0.999 | 0.794 | 0.708 | 0.737 | 0.708 |
| MPAE_onabove10 | 16.907 | 7.496 | 4.940 | 54.209 | 50.195 | 54.941 | 53.571 |

*Note:*

All evaluation metrics are based on the test sample. The training, validation, and test datasets encompass all equity options listed in the S&P 500 index from October 2019 to March 2021. The algorithms encompass various specifications of Gradient Boosted Machine (GBM, GBM2, GBM3), Black-Scholes-Merton (BSM), Cox-Ross-Rubinstein (CRR), Barone-Adesi and Whaley (BAW), and Crank-Nicolson (CN). The mean percentage absolute error (MPAE) is calculated using the following formula:

$$\frac{1}{N}\sum_{i=1}^{N} 100 * |log(\frac{\hat{y}_i}{y_i})|$$

To mitigate extreme values resulting from divergence, the CRR and CN estimates are winsorized at the 0.01% level.

Table A2: Alternative Volatility Estimates and BSM Performance

| Variable | BSM_10min_vol | BSM_1mo_vol | BSM_3mo_vol | GBM3 |
|---|---|---|---|---|
| **Entire sample** | | | | |
| MAE | 2.631 | 1.928 | 1.931 | 0.394 |
| RMSE | 10.802 | 7.221 | 7.374 | 0.883 |
| R2 | 81.334 | 91.657 | 91.300 | 99.875 |
| MPAE | 202.124 | 188.682 | 181.107 | 21.620 |
| **Out-of-the-money Options** | | | | |
| MAE | 2.911 | 1.902 | 1.933 | 0.298 |
| RMSE | 13.881 | 7.414 | 7.863 | 0.635 |
| R2 | -60.260 | 54.276 | 48.581 | 99.665 |
| MPAE | 470.153 | 463.245 | 457.521 | 31.790 |
| **At-the-money Options** | | | | |
| MAE | 2.557 | 1.928 | 1.928 | 0.414 |
| RMSE | 9.865 | 7.140 | 7.269 | 0.922 |
| R2 | 69.217 | 83.871 | 83.283 | 99.731 |
| MPAE | 152.217 | 136.852 | 128.619 | 20.205 |
| **In-the-money Options** | | | | |
| MAE | 3.029 | 2.100 | 2.000 | 0.416 |
| RMSE | 14.170 | 8.225 | 7.249 | 1.041 |
| R2 | 98.190 | 99.390 | 99.526 | 99.990 |
| MPAE | 9.869 | 7.530 | 7.556 | 2.488 |

*Note:*

The results show performance metrics by moneyness and different historical volatility estimates. A 10 million random draw from the original S&P 500 equity options is used. We remove zero price predictions from the Black-Scholes estimate for proper MPAE calculations. Moneyness is defined by +/- 10% of moneyness ratio (S/X).