What If Option Closing Prices Were Trustworthy? *

Alejandro Lopez-Lira[1], Mahendrajah Nimalendran[1], and Matthew Son[1]

[1]University of Florida

March 21, 2025

Abstract

Closing prices for equity options lack a robust mechanism similar to stocks, leading to price discovery and reliability challenges. Reporting practices rely on last trading prices and closing quotes, which are known to be unreliable due to information staleness, consolidated trades, and potential manipulation. We propose a counterfactual benchmark approach to stale prices that leverages economic models, big data, and a machine-learning (ML) algorithm, which consistently outperforms traditional models. We found the absence of robust closing option prices leads to an aggregate mispricing of X$ per day. We uncover two significant implications of ML on the options pricing literature: first, Black-Scholes input variables possess significant information content to explain option price variations. Second, time-to-maturity plays a more critical role than previously thought, surpassing the importance of volatility.

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

*We thank Narasimhan Jegadeesh for helpful comments and suggestions. alejandro.lopez-lira@warrington.ufl.edu

Daily closing prices of financial securities are one of the most fundamental data points for researchers and practitioners. The NYSE and NASDAQ use a closing auction mechanism to provide a reliable pricing mechanism and avoid last-minute consolidated trading and manipulative schemes. However, equity options lack such robust mechanisms. With options available for thousands of underlying stocks, many lack sufficient transactions to guarantee effective price discovery at market close. Consequently, options exchanges, funds, brokerages, and researchers must resort to using end-of-day mid-quotes or last trades as proxies for closing prices. To address this issue, we leverage the intrinsic relationship between options and their underlying stocks. We propose an easy-to-compute counterfactual price benchmark using a machine-learning algorithm enhanced with economic priors. This approach utilizes the underlying stock's closing price and other relevant stock characteristics as inputs to determine reliable closing prices for options.

Due to the unreliability of closing option prices, brokerages employ varying methods for reporting. Some, like Robinhood, use the rounded-up midquote of the last bid and ask, while others, such as Interactive Brokers, rely on closing prices generated by the Options Clearing Corporation (OCC).[1] The Securities and Exchange Commission (SEC) has not issued explicit regulations for reporting the value of options holdings, leaving mutual funds to typically use the last transaction price for both reporting and internal purposes. In academic research, end-of-day mid-quotes are often used for daily option returns e.g., Coval and Shumway (2001) and Muravyev and Pearson (2020)). OptionMetrics takes a different approach, using 3:59 PM prices to estimate volatility surfaces, claiming that closing prices are too noisy. Despite these varied practices, there is a scarcity of discourse regarding the reliability of closing option prices and current industry standards.

Using the last transaction as a proxy for the closing price can have severe implications. For infrequently traded options, this method can result in stale information. A large frac-

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

tion of options experience their last transaction well before 3 PM while underlying stock prices continue to be actively traded until the closing auction, resulting in a discrepancy in price discovery. Similar to the equity market, options transactions exhibit U-shaped intraday patterns where trading volume is concentrated near the close. This large volume of trading can lead to large variability in price and returns (e.g., Admati and Pfleiderer (1988) and Wood et al. (1985)) with considerable quote changes, resulting in noisy quotes at the close. Similarly, using the end-of-day mid-quote as a benchmark is also problematic. It is common to observe placeholder-like quotes at market close that are unlikely to be traded. Such inconsistencies are unsurprising, as market makers do not expect these quotes to be fulfilled. Consequently, most options data vendors avoid using 4 PM quotes for calculating implied volatility, as it generates an unreliable volatility surface.[2] Furthermore, options have significantly larger spread and sluggish quote updates compared to the underlying stocks, especially for out-of-the-money (OTM) and in-the-money (ITM) options. Hence, we are left without a robust and reliable benchmark.

In this paper, we propose a counterfactual benchmark closing option price that aligns with the underlying stock's closing auction price using a machine learning model. We treat equity options as direct *derivatives* of the underlying securities, assuming that the true spot option price aligns with the spot stock price. Our analysis begins with a comprehensive evaluation of both American and European option pricing models and various machine learning (ML) models. For the classic option models, we include Black-Scholes-Merton (BSM), Cox-Ross-Rubinstein (CRR), and other standard American options engines using finite difference methods (e.g., Barrone-Adessi-Whaley (BAW), Crank-Nicolson (CN)). For the machine-learning models, we use Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGB), Distributional Random Forest (DRF), Deep Learning (NN), and Elastic Net (EN). We then assess out-of-sample estimation accuracy for each of the models by limiting the ML model inputs to the same variables used in traditional models (such as BSM/CRR/BAW) and

[2]Data vendors provide implied volatility based on 3:45 PM or 3:59 PM instead.

employing random train-valid-test splits. Finally, we enhance the ML model's predictive efficiency by incorporating additional features that are not used in the traditional models such as VIX.

We find that traditional pricing models achieve moderately high accuracy with out-of-sample R^2 between 71% - 82%[3]. Yet, machine-learning models using the same inputs as the traditional models outperform all the traditional models, achieving a striking performance of 98.7% out-of-sample R^2 (for the GBM model). To further reduce the estimation convergence time and accuracy, we include additional features such as stock-to-strike (S/X) ratio and VIX[4], which only increased the out-of-sample R^2 to 99.8%. This finding leads us to conclude that structural options pricing input variables have sufficient information content to explain options prices. However, it is the functional form that traditional models fail to account for the observed option price patterns, while the ML models have the freedom to employ more complex functional forms and interdependencies among the features to price options. However, the complexity also makes it difficult to interpret the importance of the variables that influence the price. However, interpretable ML models provide metrics such as feature importance to analyze the features that are relatively important in predicting the price.

By utilizing various interpretable ML techniques, we highlight the features that are important for the ML model relative to the traditional models. We find that in the BSM model volatility is an important feature, while in the GBM ML time to maturity is relatively more important than volatility in predicting option prices. To our best knowledge, this is the first study to apply ML interpretability to enhance the understanding of options pricing models. A more detailed analysis of interpretability and feature importance is provided in Section XX.X.

In addition, we provide further motivation for our study by documenting the intraday transaction patterns of options during 2019-2021 based on the OPRA transactions data.

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

[4]We use an ETN product called 'VXX' as a proxy to obtain 1-minute level intraday information. Cao et al. (2020). We find that the VIX index can produce a considerable improvement for NN models.

Our analysis reveals that the majority of options experience infrequent trading activity near the market close. For instance, we observe that approximately 83% of options on S&P 500 stocks have fewer than five transactions on a daily basis. Further, only X% of the total transactions occur in the last 10 minutes of trading (see Figure 3). (Also include the numbers by moneyness)This highlights the limited liquidity and trading volume observed in these options during the closing period.

Moreover, as the market approaches its close, we find that the quote spreads for options tend to widen. This widening of quote spreads indicates a decrease in market liquidity and increased uncertainty in determining accurate prices as the trading day draws to a close.

The intraday transaction patterns and the widening of quote spreads are also indicative of the limitations and uncertainties surrounding closing prices for options.[Needs more detailed summary on this] The quote widens as the market gets near close [needs check, further writing needed].

It is commonly found in options research that uses end-of-day midquote to estimate option returns (e.g., Coval and Shumway (2001), Muravyev and Pearson (2020)). How much of daily options returns would be deviating from closing auctions?

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 best of our knowledge, this study represents the first attempt to investigate the price information at the close in the options market.

Second, our research contributes to the field of option pricing literature by exploring the implications of machine learning. We find that 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. To

shed light on the divergences between traditional models and machine learning models, we employ a model-agnostic interpretable machine learning approach.

Literature review

Prior literature extensively discusses the efficiency of closing auctions in the the stock market, particularly when they were initially introduced. For instance, Pagano and Schwartz (2003) studied the introduction of electronic call auctions at the market closing from Euronext Paris in 1996 and 1998. Their findings indicated that execution costs were reduced and price discovery for the overall market improved as a result. Barclay et al. (2008) argued that order consolidation leads to efficient prices based on their analysis of staggered auction implementations on NYSE and NASDAQ in 2004. Other researchers have also presented empirical evidence supporting the benefits of auction markets.

Bogousslavsky and Muravyev (2022) explored the demand and impact of closing auctions and concluded that the closing auction price is 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 of NYSE and NASDAQ call auctions at the close. Similalrly, Plante (2017) argued that order centralization would reduce transaction costs and enhance price efficiency in the corporate bond market. Budish et al. (2015) advocated for the implementation of frequent batch auctions over continuous trading due to the benefits of the auction mechanism. Muravyev and Pearson (2020) contend that equity options quotes are an inaccurate proxy for fair option value and that effective spreads have an upward bias. Barbon et al. (2021) discuss the intense trading activity at the close of the stock market triggered by options market makers' delta hedging.

Despite the extensive research in this area, discussions on price efficiency and the implementation of call auctions at the market close for the options market are rarely found in the literature and policy agenda, even though potential benefits exist. Our study makes a significant contribution to the existing literature by examining the textbook/traditional

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.

To the best of our knowledge, our study is the first to shed light on the specific issue of closing auctions in the context of equity options. By addressing this gap in the literature, we provide valuable insights and analysis that can enhance the understanding of the implications and dynamics of closing prices in the equity options market.

Data and Descriptive Statistics

Our dataset is constructed from multiple reliable 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. However, our analysis primarily relies on CBOE Trades and Quote data, which is provided by the Option 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, as well as volume data. By leveraging this dataset, we gain insights into the dynamics of options trading.

Our study covers a sample period spanning from October 2019 to March 2021, comprising a substantial volume of 1.7 billion records of transactions observed across 370 business days. To ensure focus and relevance, we limit our analysis to equity options listed on the S&P 500 throughout the entire sample period. Non-standard options, including FLEX options, are excluded from the analysis. 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 and mid-quote prices at the open and close are obtained from the New York Stock Exchange Trades and Quotes (TAQ) data. Specific trading condition variables denoted by "6" or "M" indicate the close, while "O" and "Q" represent the open.

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 prior to the options transaction time, and the volatility estimate is then annualized. This provides an ex-ante estimate of volatility that is current. In our enhanced ML model. we alos include the VIX index (Hull and ?? have shown that VIX has a significant contribution in explaining the volatility surface).

In Table 1 we provide 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 underlying stock (S), years to maturity (T), underlying volatility (σ), 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 (FANG Stocks) during our sample period. Further, the distribution of options trading volume is left-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 2 and 3 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 15% of the volume!), while the top 100 stocks account for more than 80% of the volume.

[Insert Figure 2 Here]

The average and the median transactions price per option are \$0.68 and \$2.35,options transaction prices per contract, it is observed that the first quartile and median values amount to \$6.65 and \$2.35 with a right-skewd distribution. These figures highlight the frequent occurrence of retail-sized options trading, where smaller-sized transactions are prevalent. 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 terms of their magnitude.

Options Intraday Trading Pattern

In terms of intraday trading, we observe a U-shaped pattern with trading activity concentrated and more pronounced near the market open and close, similar to Chan et al. (1995); Bergsma et al. (2020). This U-shaped pattern indicates a higher level of trading interest, liquidity, price discovery, and information efficiency during these periods.

[Insert Figure 3 Here]

Figure 3 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 prior to 3:40 PM. Furthermore, a significant number of 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 the majority of options trading occurs earlier in the trading day, contributing to the challenge of obtaining reliable prices at the close of trading.

Machine Learning Models and Estimation

Selecting the optimal machine learning algorithm for a given task poses a significant challenge due to its empirical and data-driven nature. 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.

Traditional Models

The Black-Scholes-Merton (BSM) European options model and the Cox-Ross-Reubenstein (CRR) binomial tree model have gained widespread adoption among researchers and practitioners, and we mainly use them as our benchmark/reference point. To estimate option prices accurately, we initiate our approach by ensuring fair comparisons with traditional models. Both American and European style option pricing models share common input variables, which can be concisely summarized by appropriate functions described 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 measurement of volatility requires careful consideration. The BSM model assumes constant volatility over the maturity of the option, and historical stock returns are used to estimate the volatility

input for the models. We employ realized volatility calculated over the most recent 10-minute period as a proxy for volatility in all our models. [We also use volatility estimates over longer periods as a robustness check.??]

Machine Learning Models

To facilitate the formulation of our machine-learning algorithms, we use the functional form described by Equation 2.

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

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

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_{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 $M = (f, \theta_{\text{best}})$

ML Estimations Strategy

Machine learning algorithms are not constrained by the number and type of input variables. In fact, expanding the set of input variables and incorporating additional relevant data can significantly enhance their performance. However, we intentionally impose limitations on the input variables to be identical to the ones used by the classical models for the following reasons. First, this allows us to investigate scenarios where traditional options pricing models may falter due to either 1) insufficient information contained 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, it is customary to employ a rolling or recursive sample-splitting scheme for forecasting purposes. This scheme involves shifting the training and validation samples forward to encompass more recent data, as exemplified by Gu et al. (2020). Another commonly utilized approach for sample splitting is the fixed scheme, which randomly partitions the data without replacement based on a specified split ratio. Given the nature of option pricing models, which do not rely on time-specific information but rather on the remaining lifespan of the options contract (τ), the fixed-scheme approach is particularly 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 integrity of our analysis.

To streamline the estimation process considering the multitude of hyperparameter dimensions and algorithms involved, we adopt a two-fold approach: 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.[5] This validation set serves the crucial function of fine-tuning the hyperparameter search

[5]To validate the robustness of our approach, we conducted a comparative analysis between the train-valid-

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.

Model Estimates and Performance Evaluation - 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 3 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 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 larger errors.

Interestingly, despite being designed for European-style options, the BSM model displays lower average bias when compared to American options models[6]. Alternative American options models, such as Cox-Ross-Reubenstein (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%.

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

test split and 5-fold cross-validation methods. Remarkably, both approaches exhibited similar performance outcomes, while the former approach significantly reduced the overall estimation time.

[6]Nevertheless, it is important 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.

performance across all evaluation metrics, boasting a remarkably low MAE of 1.2 and RMSE of 2.9, surpassing both the BSM model and CRR model. Our performance results highlight that structural models, including the BSM and CRR models, tend to exhibit larger estimation errors compared to the GBM model. It is worth noting that the overall out-of-sample R-squared value of 98.7% should be interpreted with caution, as R-squared metrics based on a level variable can be influenced and dominated by a certain scale. The appendix contains performance results based on the dollar scale of options.

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 a remarkable finding, particularly 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 4 Here]

The provided table 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. However, it is notable that structural

models outperform machine learning models in explaining ITM options. For instance, the Black-Scholes-Merton (BSM) and Cox-Ross-Reubenstein (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.

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 the inclusion of additional input variables to potentially enhance model performance, particularly for the best-performing algorithm, GBM. 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 (σ) 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.

Significant improvements are observed 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, as a result of the proposed modifications. In the first modification, the MAE on the test set decreased from 1.276 to 0.584 (a reduction of 54%). Further enhancement is achieved in the second modification, where the MAE decreases to 0.395 (a reduction of 69%). Similarly, the RMSE demonstrates a notable reduction of 55% and 71% in the first and second modifications, respectively.

[Insert Table 5 Here]

The above table presents a comprehensive overview of the detailed performance results. 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 has effectively ensured a balance between underfitting and 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 datase with respect to 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 10 Here]

The figure presented depicts the relationship between true option prices and the corresponding predictions generated by our four selected models. Upon careful examination, distinct patterns emerge in relation to the magnitude of prediction errors and variances exhibited by the Black-Scholes-Merton (BSM) and Cox-Ross-Reubenstein (CRR) models. Specifically, as the price level decreases, these structural models tend to demonstrate larger prediction errors and greater variability. 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 smaller variations in prediction accuracy across different price levels.

Interpretable Machine Learning

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. 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 in identifying the most influential factors that contribute to accurate predictions, enabling domain experts to validate or challenge 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 risks.

The determination of feature importance is typically based on the model's internal mechanisms or through specific algorithms designed for this purpose. Various approaches exist, such as permutation importance, Gini importance, or coefficients derived from regularization techniques like Lasso or Ridge regression. These methods assign scores or rankings to features, reflecting their significance in influencing the model's output.

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

[Insert Figure 4 Here]

As illustrated in the above figure, all of the machine learning algorithms unanimously identify the stock price (S) and strike price (K) of options as the most influential variables for predicting option prices. This unanimous 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 slightly less importance on these variables, although they remain crucial for accurate prediction.

As demonstrated by Hull and White (1987), in situations involving stochastic volatility, particularly when there is a correlation between stock price and volatility, the BSM model exhibits deviations from its predicted prices. Notably, they found these deviations tend to

increase as the time to maturity of the option lengthens. The relatively diminished significance attributed to volatility in our analysis can be attributed to the adoption of historical volatility (σ) as a proxy. Instead, we observe that the time to maturity (T) variable encompasses more relevant information pertaining to 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 delve into the examination of 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.[7]

Figure 9, presented below, showcases the feature importance of the BSM model:

[Insert Figure 9 Here]

Drawing a comparison with Figure 5 presented earlier, we observe both similarities and dissimilarities in the feature importance rankings. Notably, both models identify the strike price (K), stock price (S), and time to maturity as the most significant variables in determining option pricing accuracy. Conversely, neither model attributes significant importance to the dividend yield ratio as it does not contribute substantially to improving model performance. However, a notable disparity arises in the treatment of underlying volatility (σ) and risk-free rate (r). The GBM model places considerably less emphasis on volatility, with it accounting for less than 10% of the most important variable. In contrast, the BSM model assigns a greater level of importance to volatility, constituting more than 50% of the most important variable.

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

[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. We examined two models and found that the coefficient estimates from the GBM model were consistently smaller, indicating lower susceptibility to errors across all dimensions.

Both models showed larger deviations for out-of-the-money options and smaller deviations for in-the-money options compared to 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 compared to short and very long maturities.

However, there are differences between the models. The Black-Scholes-Merton (BSM) model tends to make more 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.

Estimating Counterfactual Closing Option Price

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

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

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, specifically the last transaction price and midquote at 4 PM.

Price Deviations from ML Benchmark

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

[Insert Table 8 Here]

In 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 greater variance.

In Panel B, we investigate the deviations compared to the last traded price. 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 compared to 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 options of large firms display less variation, with a range of 31 compared to 50 for small firms.

Note that the number of last transactions is lower due to the presence of 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 aforementioned findings.

[Insert Table 9 Here]

Table 9 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 tend to decrease, leading to larger percentage deviations when compared 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. In comparison to the deviations calculated using the 4 PM quote, which is 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.

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:

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

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.

[8] Price deviations in options can be attributed to various factors, including the presence of 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

[8]This paragraph needs to be amended after the regression results.

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 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.

To examine the influence of trading volume and its relationship with price deviations in options, we analyze two separate regressions. 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 transaction volume during the day as an independent variable to assess its association with price deviations. We aim to determine whether lower trading activity throughout the day is linked to a higher likelihood of stale quotes and subsequently larger price deviations.

In the second regression, we introduce the volume of trades specifically within the last minute of trading. This variable allows us to examine the influence of noise trades occurring 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]

[9] The estimation results are presented in Table **??**. In order to illustrate the extent of

[9]This paragraph needs to be amended after the regression results.

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 in support of the trade noise hypothesis.

Analyzing the deviations based on the 4 PM midquote, we find that put options exhibit lower deviations compared to call options. Furthermore, options with 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.

In relation to the last traded prices, our analysis does not yield statistically significant evidence supporting the influence of firm size on deviations. Nevertheless, we observe a noteworthy trend where both the top 15 and top 50 most traded underlying companies consistently exhibit significantly lower deviations. This observation suggests a potential correlation between deviations and trading activities, indicating that the level of trading volume and market liquidity may play a role in reducing pricing deviations at the close.

The positive relation between bid-ask spread and deviation shows that options with larger spreads are likely to have larger deviations. This relation holds after controlling the moneyness of options. Hence, this implies that the larger the spread, the noisier the midquote of options. In association with volume, this raises concerns regarding the validity of using close midquote of options: deviations are significantly greater for the options that are actively traded and have larger spreads.

Options are characterized by large bid-ask spreads, indicating a significant difference between the prices at which market participants are willing to buy and sell these instruments. Furthermore, compared to the underlying equities, options exhibit slower updates in their quote information. These distinctive features of options trading have prompted recent studies to question the reliability and timeliness of using midquotes as a proxy for capturing the true price process (Muravyev and Pearson (2020)).

[Further work needed on price estimate and bid-ask bound: are they away or within the bid-ask spread?] Prior writing: Due to the large spread, the counterfactual option close price lies within most options' bid-ask spread. Our study shows that price deviation at the close is pronounced for the options that are actively traded (i.e.~, large day volumes, and shorter expirations), and has larger spreads. It raises concerns about using close midquote for valuing closing options.

[Onother work needed: overnight options return calculation, based on the benchmarks]

Prior writing: In the next step, we aim to study how fast price deviations revert back overnight and expect them to be slower than the stocks due to the noisiness and sluggishness of option quotes.

Conclusion

This part will be written based on the introduction...

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.

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.

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.

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)*.

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.

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

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.

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

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*.

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.