Inferring Trade Directions in Options via Machine Learning *

Si Gao† Xiaoxiao Tang‡ Guofu Zhou§ Zhaoque (Chosen) Zhou¶

Current Version: March 2025

*

†Washington University in St. Louis, John M. Olin Business School, gao.si@wustl.edu
‡Jindal School of Management, University of Texas at Dallas, xiaoxiao.tang@utdallas.edu
§Washington University in St. Louis, John M. Olin Business School, zhou@wustl.edu.
¶Washington University in St. Louis, John M. Olin Business School, zhaoque@wustl.edu.

Inferring Trade Directions in Options via Machine Learning

Abstract

We develop GS-LASSO, a novel machine learning approach integrating XGBoost, SHAP, and LASSO, to classify trade direction in the evolving options market. Structural changes, particularly the rise of retail options trading and price improvement auctions, have rendered traditional classification methods ineffective, with accuracy falling below 60%. Using a proprietary dataset from the Philadelphia Stock Exchange (PHLX), our GS-LASSO model achieves 75.94% accuracy, outperforming conventional methods by approximately 20%. This improvement enhances market microstructure estimates, provides new insights into wholesalers' profitability, and supports recent SEC regulatory initiatives aimed at increasing competition in retail stock trading through order-by-order auctions.

Keywords: Stock options, ETF options, Midpoint rule, Machine Learning, SLAN, XGBoost.

1 Introduction

Accurately identifying trade direction is essential for many research questions in financial markets, and various algorithms have been developed to classify trades with high accuracy. However, the options market has undergone dramatic structural changes in recent years, particularly with the surge in retail options trading following the rise of zero-commission trading since 2019. Unlike in the past, a significant portion of retail option orders is now routed to wholesalers through Payment for Order Flow (PFOF), where price improvement auctions determine trade execution. This shift has fundamentally altered how trades are processed, making traditional trade classification methods less effective. Using a proprietary dataset from the Philadelphia Stock Exchange (PHLX), we find that existing methods now achieve no more than 60% accuracy, only slightly better than random guessing. These changes highlight the urgent need for a new approach to trade direction classification that reflects the evolving structure of the options market.

In this paper, we introduce GS-LASSO, a novel machine learning framework that integrates XGBoost, SHapley Additive Explanations (SHAP), and LASSO to infer trade directions with greater accuracy. Built entirely on public data, GS-LASSO achieves an out-of-sample accuracy of 75.94%, consistently outperforming traditional classification rules by approximately 20%. This significant improvement enhances the precision of market microstructure measurements and provides new empirical insights into wholesalers' profitability and retail trading behavior. Furthermore, our findings contribute to ongoing regulatory discussions, offering empirical support for the SEC's recent initiatives to increase competition in retail stock markets through order-by-order auctions.

Our study utilizes a proprietary options dataset from the Philadelphia Stock Exchange (PHLX), which ranked second in market share for U.S. equity options in 2023. Our dataset covers January 1, 2024, to December 31, 2024, and includes 46.68 million detailed auction messages, transaction-level data from OPRA, and tick-by-tick trade and quote data, allowing us to analyze retail-driven

price improvement auctions and their impact on market microstructure. Additionally, we construct the Single-Lag Auction Non-ISO Trade (SLAN) dataset, which captures retail-driven trades and accounts for 63.10% of PHLX's options volume, making our dataset representative of broader market activity. We find that the widely-applied midpoint rule is ineffective for SLAN, achieving only 56.51% accuracy.[1] This misclassification distorts key market metrics, including transaction costs, effective spreads, and price improvements, leading to biased estimates of order imbalances and trading behavior.

The low accuracy of the midpoint rule stems from wholesalers frequently offering price improvements exceeding half the quoted spread during auctions. We are the first to show that over 40% of SLAN trades receive such improvements, suggesting that wholesalers are highly motivated to capture negative quoted spread revenue from certain investors.[2] This finding challenges traditional theoretical models (Garman (1976), Amihud and Mendelson (1980), Ho and Stoll (1981)) that assume market makers earn positive spread returns to compensate for transaction costs, inventory risk, and information asymmetry. However, while wholesalers may incur losses on the spread for some trades, we show that they can profit from holding these positions, indicating their ability to differentiate between informed and uninformed trades. Wholesalers provide substantial price improvements for uninformed trades, expecting to recoup losses through strategic position management, while for informed trades, they offer limited price improvement and actively close or hedge positions before potential losses exceed the spread revenue.

Recognizing that price improvement depends on the true trade direction and assuming wholesalers determine it primarily based on order and publicly available data, we develop machine learning models to classify SLAN trades using only publicly accessible information. Our baseline XGBoost model achieves an out-of-sample accuracy of 73.11%. To enhance both accuracy and interpretability, we apply SHAP (Lundberg and Lee (2017)) to explain the non-linear model and

[1] The midpoint rule classifies a trade as a buy if the price exceeds the quoted midpoint and as a sell if it falls below.

[2] When traders receive a price improvement exceeding half the quoted spread, they can buy below the midpoint and sell above it, while wholesalers, as the counterparty, sell below the midpoint and buy above it, effectively paying the spread on more than 40% of trades.

introduce a novel feature selection method using LASSO on SHAP values. This GS-LASSO model further improves accuracy to 75.94%, consistently outperforming two midpoint rule variants—one based on the quoted midpoint and the other on implied volatility—by approximately 20%.

The GS-LASSO model effectively identifies the most relevant features for trade classification while shrinking less significant ones. It selects 36 out of 50 features, including dollar-quoted spread, trade size, option return over the past trading period, short-term option volatility, and midpoint rule classification. The failure of the midpoint rule is largely attributed to its omission of these key variables. Our model demonstrates strong performance across various XGBoost settings and consistently outperforms traditional logistic regression models in accuracy.

The improved accuracy in trade direction classification enhances the reliability of market microstructure estimates. For instance, while the midpoint rule estimates the volume-weighted average price improvement at 19.05% of the quoted spread, our GS-LASSO model revises this figure to 36.49%. Given that the true trade direction implies a price improvement of 32.21%, our GS-LASSO estimate is significantly closer to the actual value, highlighting the limitations of traditional classification methods.

Additionally, our study provides the first empirical estimates of wholesalers' profits on SLAN trades, showing earnings ranging from 36.87 to 81.98 cents per contract, depending on how effectively they manage their positions. These margins support the zero-commission business model, despite wholesalers paying only 0.28 to 0.49 cents per contract in PFOF (Hendershott et al. (2022)). Our findings also reveal that wholesalers strategically adjust price improvements based on trade informativeness, offering higher price improvements for uninformed trades while limiting concessions for informed trades. When wholesalers extend price improvements exceeding half the quoted spread to acquire positions, they profit from briefly holding these uninformed trades, explaining their willingness to offer substantial price enhancements.

Finally, our paper provides important regulatory insights. The SEC has recently proposed reforms to increase competition for retail stock orders by requiring market makers to conduct

order-by-order auctions.[3] Our study highlights that similar price improvement auctions have been successfully implemented in the options market since the early 2000s, where all trades are executed on exchanges. This auction-based structure enhances allocative efficiency, as suggested by theoretical models (Ernst et al. (2022)). Our empirical findings support this view, suggesting that introducing similar mechanisms in the stock market could lead to significant price improvements for retail investors.

Our study contributes to several strands of literature. First, it relates to the growing body of research applying machine learning tools in finance. Studies such as Light et al. (2017), Chinco et al. (2019), Freyberger et al. (2020), Gu et al. (2020), Kozak et al. (2020), Avramov et al. (2023), Chen et al. (2024), and Han et al. (2024) have leveraged machine learning for asset pricing, primarily to explain cross-sectional returns. More recently, Bogousslavsky et al. (2024) applied XGBoost to detect informed trading intensity, while Bartlett and O'Hara (2024) used it to uncover hidden liquidity. In contrast to these studies, our paper is the first to apply machine learning techniques to classify trade direction. Complementing the approaches in Bogousslavsky et al. (2024) and Bartlett and O'Hara (2024), which employ XGBoost for market microstructure analysis, we enhance the model by incorporating LASSO for feature selection through SHAP. This methodological refinement improves traditional XGBoost classification, offering a more interpretable and effective framework for trade direction identification.

Second, our study contributes to the growing literature on price improvement auctions in option markets. Bryzgalova et al. (2023) introduce SLAN orders and suggest that SLAN trading serves as a proxy for retail option activity. Ernst and Spatt (2024) examine option trade internalization through price improvement auctions and designated market maker routing, arguing that internalization in the options market remains imperfectly competitive. In contrast, Hendershott et al. (2022) develop a stylized model showing that auction-based mechanisms foster greater competition than non-auction settings, while Anand and Muravyev (2024) find that restricting

[3]For further information regarding the rule, refer to the SEC website: https://www.sec.gov/newsroom/press-releases/2022-225.

auction usage through regulatory changes can lead to narrower quoted spreads and reduced quote competition. Our study is the first to provide empirical evidence that wholesalers offer substantial price improvements in the options market and that auction-based trades are highly profitable. These findings highlight the role of price improvement auctions in shaping market dynamics and further inform ongoing debates on auction-based trading mechanisms.

The remainder of the paper is organized as follows. Section 2 describes the dataset used in our analysis. Section 3 provides robust evidence on the low accuracy of the midpoint rule in classifying trade direction. Section 4 introduces our XGBoost-based machine learning models and the GS-LASSO framework for improved classification. Section 5 investigates the profitability of SLAN trades and examines how wholesalers adjust price improvements based on trade characteristics. Finally, Section 6 concludes.

2 Data

Our study mainly uses the extracted auction messages from the Philadelphia Stock Exchange (PHLX), which ranked second in market share for equity options among the 17 U.S. options exchanges in 2023.[4] We first collect historical data from the PHLX orders feed, which provides the state of both simple and complex orders on the PHLX book from January 1, 2024, to December 31, 2024. This dataset includes nanosecond-level details on orders added, modified, canceled, filled, and auctioned across all tradable option series on PHLX. In this paper, we extract all simple order price improvement auction messages, capturing key details such as order direction (buy or sell), option contract specifications, order timestamps, auction prices, and auction volumes. These orders are then matched with transaction-level data from the Options Price Reporting Authority (OPRA), obtained via ThetaData, using timestamps and option contract identifiers.

We then construct the Single-Lag Auction Non-ISO Trade (SLAN) dataset, where trades are

[4]In 2023, PHLX held an 11.29% market share in equity options, the second highest among the 17 U.S. options exchanges, trailing only ARCA, which had a 12.18% share. Data source from OCC: https://www.theocc.com/market-data/market-data-reports/volume-and-open-interest/volume-by-exchange

identified by the SLAN trade condition in the OPRA trade dataset, aligning with the retail options trade classification in Bryzgalova et al. (2023). Our dataset collects 46.68 million SLAN trades executed on PHLX with their true trade direction. In 2024, SLAN trades accounted for 63.10% of PHLX's total options volume, aligning with Bryzgalova et al. (2023), which estimated that SLAN trades comprised over 60% of the total market volume in early 2020. This suggests that our dataset is representative of the overall market. To enhance our analysis, we further match the SLAN trade dataset with tick-by-tick OPRA option quote data, as well as tick-by-tick underlying trade and quote data from TAQ.

Table 1 presents a range of daily trade statistics for the SLAN dataset of one year from January 1, 2024 to December 31, 2024. We apply several filters to our dataset: (1) remove observations with trade prices outside the quoted spread (i.e., prices above the ask or below the bid); (2) exclude observations with an ask price that is less than or equal to the bid price; (3) discard observations with non-positive trade prices; and (4) filter out observations with trade sizes that are less than or equal to zero or exceed 250,000.[5] These exclusions account for only about 0.93% of the total dataset. The average daily trade volume is 1.57 million contracts, with a range spanning from 0.5 million to 2.38 million. Notably, the peak volume exceeds 4 times the lowest recorded value.Trading activity varied throughout the year, with the market showing noticeable differences in dynamism from one day to the next.

Our classification of at-the-money (ATM), in-the-money (ITM), and out-of-the-money (OTM) options is based on precise trade-time quotes rather than end-of-day quotes. According to our criteria, ATM options, defined as those with moneyness between 0.95 and 1.05, are the most prevalent, accounting for 61.54% of total trading volume, followed by OTM options at 24.55% and ITM options at 13.88%. Moreover, in our sample, call options dominate both in trading volume and number of trades, with a call-option volume ratio of 65.00% and a similar proportion in number of trades. Additionally, approximately 50.61% of total SLAN trades originate from buy orders,

[5]The 250,000 is the highest position limit set by NASDAQ PHLX. Details refer to Section 13(g) on https://listingcenter.nasdaq.com/rulebook/phlx/rules/.

indicating that buy and sell volumes are nearly balanced in these auctions. Furthermore, 94.64% of SLAN trades in PHLX are written on stocks and ETFs, while the remaining 5.36% are based on the NASDAQ index, ETPs, ADRs, and other instruments. Stock options account for 68.57% of the total volume, while ETF options comprise 26.07%, reflecting the growing popularity of ETF options in recent years.

Interestingly, ETF options and stock options exhibit distinct trading patterns in our sample. First, the SLAN trade size for ETF options is significantly larger than that of stock options. On average, stock options account for 150,553 trades per day, representing approximately 80% of overall daily trades, while ETF options total around 34691 trades, or roughly 20%. Therefore, the average trade size for ETF options is 12.68 contracts, exceeding the 7.57 contract average for stock options. Across all stock and ETF SLAN trades, the average trade size is 8.49 contracts per trade, with 47.3% of transactions occurring at the minimum size of one contract. This pattern corroborates NYSE's operational definition where no more than 10-contract trades predominantly represent retail participation[6], suggesting retail investors constitute the majority of SLAN activity.

Second, the quoted spread for ETF options is significantly tighter than that for stock options. In our sample, the overall percentage quoted spread is 0.1509, slightly lower than the 0.172 reported by Muravyev and Pearson (2020), which analyzed option data for S&P 500 stocks from 2004 to 2015. This suggests that option market development has led to tighter spreads over time. However, when examining stock and ETF options separately, ETF options, a relatively newer segment of the market, exhibit a tighter percentage quoted spread of 0.1151, compared to the daily average percentage quoted spread of 0.1648 for stock options.

As wholesalers commit to providing price improvement to SLAN traders, we measure the price

[6] https://www.nyse.com/data-insights/trends-in-options-trading

improvement for individual trade as:

$$PI = \begin{cases} \frac{P_{Ask} - P_{Actual}}{\text{Quoted Spread}} & \text{if buy trades} \\ \frac{P_{Actual} - P_{Bid}}{\text{Quoted Spread}} & \text{if sell trades.} \end{cases} \tag{1}$$

Table 2 presents price improvement (PI) statistics derived using the true trade direction (TTD) from our proprietary PHLX dataset. Panel A shows that the percentage price improvement is relatively stable across different trade categories, with buy and sell trades receiving 43.5% and 43.0% PI, respectively. Moreover, buy trades achieve significantly higher PIs than sell trades (t-stat = 66.59), suggesting that wholesalers may favor option buyers to collect premiums from these transactions. On the other hand, calls exhibit a 43.1% PI, which is significantly lower than the 43.5% PI observed for puts with t-stat of -53.70.

Panel B shows that the distribution of dollar price improvements is right-skewed, with many trades receiving only modest improvements in dollar terms. Notably, more than half of the transactions obtain price improvements of no more than two cents. This reflects the constraints imposed by the one-cent minimum tick size in options markets, which may prevent wholesalers from offering greater price improvements, as sub-penny pricing, common in the stock market, is not permitted in the options market. These findings suggest that while percentage price improvements can appear large, they are often driven by tight quoted spreads, potentially creating a misleading impression of low liquidity when measured by the percentage quoted spread. For instance, for out-of-the-money (OTM) options quoted at [0, 0.05], the percentage quoted spread can reach 200%.

3 Failure of Midpoint Rule

In this section, we provide robust evidence demonstrating that the low accuracy rate of the midpoint rule persists across various intraday time intervals and applies to both stock and ETF options.

Additionally, we show that this low accuracy rate is not driven by quote mismatching. Instead, in the absence of quote mismatching, misclassification under the midpoint rule primarily results from large price improvement offered by wholesalers, which exceeds half of the quoted spread.

3.1 Low Midpoint Rule Accuracy Rate

To classify option trades as buys or sells, researchers often apply stock trade classification rules, including the midpoint rule and the method proposed by Muravyev (2016). Savickas and Wilson (2003) utilize a proprietary CBOE dataset in 1995 to evaluate four stock trade classification rules: the midpoint rule, the Lee and Ready (1991) method, the Ellis et al. (2000) method, and the tick rule. Their findings confirm that the midpoint rule, which classifies trades executed above (below) the midpoint as "buy" ("sell") trades, achieves the highest accuracy at 83%. Muravyev (2016) further refine this approach by proposing that trades at the midpoint be classified based on the quote rule of the exchange where the trade occurred. Bryzgalova et al. (2023) confirm that the ticker-level order imbalance computed using these methods exhibits a high correlation exceeding 94%.

However, the structure of the options market has undergone significant changes since the 2000s. Price improvement auctions, such as CBOE's AIM, ISE's PIM, and BOX's PIP, were introduced in the early 2000s to enhance trade execution quality. Additionally, the Options Order Protection and Locked/Crossed Market Plan, which functions similarly to Reg NMS Rule 611 (the Order Protection Rule) in the stock market, was officially implemented on August 31, 2009. This plan ensures that options investors can execute trades at the NBBO or better prices. Due to these structural changes, the accuracy of the midpoint rule in classifying option trades may no longer remain at high levels. In this section, we use our proprietary PHLX dataset to confirm that the accuracy of the midpoint rule has significantly declined, dropping to no more than 60% in 2024. Furthermore, this low accuracy remains consistent across intraday time intervals and applies to both stock and ETF options.

We primarily focus on the accuracy of the midpoint rule for two reasons. First, all option classification rules are variations of the midpoint rule, and Bryzgalova et al. (2023) demonstrates that the order imbalances computed using these methods are highly correlated. If the midpoint rule exhibits low accuracy, other methods are likely to be similarly inaccurate. Second, OPRA currently distributes only the NBBO and trade messages to data vendors.[7] As a result, data vendors receiving data from OPRA do not have access to exchange-level quote data, which is essential for the method proposed by Muravyev (2016).

As seen in Figure 1, the midpoint rule exhibits low accuracy at the daily level. Figure 2 further illustrates the intraday behavior of midpoint rule accuracy for stock and ETF options, revealing a decline in accuracy as the trading session progresses. Moreover, the stock options exhibit higher accuracy than ETF options. Given that stock options account for approximately 80% of the total number of trade in SLAN, the aggregate midpoint rule accuracy largely mirrors the trajectory of stock options. Furthermore, option trading activity is not evenly distributed throughout the day, with significantly higher volumes during the first hour after market open. The accuracy of the midpoint rule follows a similar pattern, with greater accuracy observed during periods of heightened trading intensity. This trend may be attributed to variations in trading intensity dynamics across trading sessions. During peak trading intervals (the first two half-hour intervals), increased trading activity enhances natural counterparty matching, enabling wholesalers to fulfill order flow with minimal price improvement (PI). Conversely, during off-peak periods, lower trading intensity incentivizes wholesalers to offer greater PI to attract order flow. In such cases, when PI exceeds 50%, it leads to misclassification under the midpoint rule.

To determine whether intraday midpoint rule accuracy is driven by trading intensity, we regress intraday accuracy on the volume-weighted average spread, log volume, and a VIX proxy represented by the VXX ETF midpoint quote price, while controlling for intraday fixed effects. Table 3 shows that, overall, higher option volume and an increase in VIX are associated with

[7]For more details, see Section 5.2 of the OPRA Plan: https://cdn.opraplan.com/documents/OPRA_Plan.pdf.

improved accuracy of the midpoint rule. However, when the sample is split into stock options and ETF options, distinct patterns emerge. For stock options, a higher VIX corresponds to reduced accuracy, and the quoted spread shows no statistically significant effect. In contrast, for ETF options, both a higher VIX and a wider quoted spread enhance accuracy. Despite these differences, both subsamples consistently indicate that larger option volume leads to higher accuracy. This observation aligns with our intraday pattern in Figure 2, which shows elevated accuracy during periods of peak trading volume.

3.2 Limited Quote Mismatching Issue

The midpoint rule classification relies heavily on the midpoint price, which is calculated using the bid and ask prices. Inaccurate or mismatched quote prices can result in poor performance of the midpoint rule. Holden et al. (2023) argue that quote mismatches reduce the accuracy of the midpoint rule in determining stock trade direction. After adjustments, the trade classification accuracy of Lee and Ready (1991) increases from 86% to 92% in the stock market. In this section, we present three pieces of evidence confirming that there is no significant quote mismatching in our dataset. Therefore, the low accuracy rate of the midpoint rule is primarily due to the substantial price improvement offered by wholesalers.

After a PHLX member submits a SLAN order on behalf of their customers, an auction message is recorded in the PHLX TotalView dataset and broadcast to subscribers of PHLX's TOPO Plus data feed, as well as market makers. The auction lasts for 100 milliseconds, during which all participants can submit responses. These responses remain confidential and are not visible to others. Once the auction concludes, the recorded trade is sent to OPRA.[8] We match SLAN trades recorded on OPRA with SLAN auction messages from the PHLX TotalView dataset. As a result, the expected timestamp difference between OPRA and PHLX records should be approximately 100 milliseconds.

[8]For more details about the PHLX auction, refer to https://www.nasdaqtrader.com/content/phlx/pixlfaqs.pdf.

Figure 3 presents the time series of the timestamp differences. On average, the daily median timestamp difference is 99.54 milliseconds, implying an average latency of 0.46 milliseconds (=100 - 99.54) between OPRA and PHLX. Additionally, the average 5th percentile timestamp difference is 99.08 milliseconds, indicating that fewer than 5% of SLAN trades experience a latency greater than 0.92 milliseconds (=100 - 99.08). A median timestamp difference of less than 100 milliseconds suggests a slight delay in recording SLAN auction messages in the PHLX TotalView dataset. This is supported by the spike observed on November 11, 2024, when Nasdaq upgraded its NY11 data center in Carteret, NJ and opened access.[9] Consequently, the timestamp difference increased from 99.53 to 99.58 milliseconds after that date, indicating a decrease in latency from 0.47 to 0.42 milliseconds. We note that OPRA timestamps are recorded to the millisecond, therefore, delays of less than one millisecond should not significantly affect the accuracy of the midpoint rule.

If the low accuracy is primarily due to quote mismatches, using quote data from other time points to compute the midpoint price should result in a significant increase in accuracy. However, Table 4 shows that the midpoint rule consistently exhibits low accuracy even when using quote data from various time points. This suggests that the failure of the midpoint rule is not solely due to a mismatch in quote timing.

Panel A of Table 4 reports the average accuracy of the midpoint rule for stock and ETF options using quotes from various time points: one second before the trade, 100 milliseconds before the trade, last quote before the trade, the contemporaneous OPRA quote, next quote after the trade, and one second after the trade. If quote mismatches were the primary cause of the low accuracy, we would expect a significant improvement when using quotes around the trade. However, the results indicate that accuracy remains consistently low across all time points, with only minor fluctuations. The highest accuracy is observed for the contemporaneous quote (55.82%), followed closely by the quote immediately after the trade (55.83%). These figures are far below the over 80% accuracy

[9]For more details, refer to https://listingcenter.nasdaq.com/assets/rulebook/nasdaq/filings/SR-NASDAQ-2024-056.pdf.

typically observed in the stock market, indicating that the midpoint rule's low effectiveness in the options market cannot be attributed solely to quote mismatches.

Finally, if quote mismatches were the primary cause of the midpoint rule's low accuracy, trades with slower quote updates should exhibit higher accuracy. To test this, we split the sample into two equal groups based on the time difference between the last quote and the trade on OPRA, where a larger time difference indicates slower quote updates. Contrary to our expectations, the group with the larger time difference showed significantly lower accuracy for the midpoint rule, which rejects the quote mismatch hypothesis.

To sum up, advanced technology at the exchange has resulted in minimal latency between the exchange and OPRA. In addition, our analysis, matching trades with quotes at various time points and conducting a subsample analysis, consistently rejects the quote mismatch hypothesis. This indicates that the low accuracy of the midpoint rule is likely due to wholesalers increasing the frequency of large price improvements among SLAN trades.

3.3 Classification Bias for Midpoint Rule

In our SLAN dataset, investors show a slight preference for buying options, with an average daily buy volume ratio of 50.64%. However, the midpoint rule exhibits significant classification bias, estimating a daily buy volume ratio of only 44.69%. This bias may lead research based on the midpoint rule to erroneously conclude that retail investors predominantly sell options. In contrast, our proposed GS-LASSO model substantially reduces this bias, yielding an average estimated buy volume ratio of 50.20%, closely aligning with the true value of 50.64%.

4 Machine Learning Identification Model

In this section, to identify the True Trading Direction (TTD), we introduce SHAP-LASSO as a feature selection method for machine learning models and propose an XGBoost model utilizing

the selected features. We call our model as GS-LASSO model. We begin by presenting SHAP (SHapley Additive exPlanations) and explaining how it can be integrated with LASSO regularization to identify the most influential features in our dataset. Next, we discuss the principles and advantages of the XGBoost algorithm, emphasizing its suitability for high-frequency financial data and how careful parameter tuning can address non-linearities. Finally, we detail our model setup, compare the performance of 18 different model configurations, and demonstrate that one-week rolling GS-LASSO model yields the highest out-of-sample accuracy.

4.1 Variable Selection via SHAP-LASSO

SHAP (SHapley Additive exPlanations), proposed by Lundberg and Lee (2017) and cited over 30,000 times, is widely used to interpret machine learning model outputs by assigning importance values to each feature for every observation. Since most machine learning methods do not fully address feature selection, we introduce the SHAP-LASSO approach to objectively identify features with the highest contributions. By decomposing predictions from non-linear models into a linear additive sum of individual feature contributions, SHAP enables the use of linear feature selection techniques, such as LASSO, to efficiently identify the most influential features.

In computer science, numerous methods have been proposed to interpret complex models. Lundberg and Lee (2017) introduce a unified framework that satisfies key interpretability properties and propose SHAP (SHapley Additive exPlanations) values as a standardized measure of feature importance.

Let $f(\mathbf{x})$ be a model with N features $\{x_1, x_2, \ldots, x_N\}$. For a given feature x_i, the Shapley value $\phi_i^k(f)$ quantifies its marginal contribution to a specific prediction $\hat{f}(\mathbf{x})$ for the k-th observation by averaging over all possible subsets of the other features. Formally,

$$\phi_i^k(f) = \sum_{S \subseteq \{1,\ldots,N\} \setminus \{i\}} \frac{|S|!\,(N-|S|-1)!}{N!} \Big[\hat{f}\big(S \cup \{x_i\}\big) - \hat{f}(S)\Big], \tag{2}$$

where S denotes any subset of features excluding x_i, $|S|$ is the number of features in S, and $\hat{f}(S)$ represents the model's prediction based solely on the features in S. By evaluating how the prediction changes when x_i is added to various subsets, SHAP effectively captures the "fair share" of the model output attributable to that feature.

SHAP values define an additive feature attribution method, allowing the transformation of nonlinear model behavior into a linear representation. In classification tasks like ours, where the output is interpreted as a predicted probability for k-th observation $p_{(k)} = P(\hat{y} = 1 \mid \mathbf{x}_{(k)})$, results are often expressed in log-odds form:

$$\text{logit}(p_{(k)}) \;=\; \beta_0 + \sum_j \phi_j^{(k)}, \tag{3}$$

with β_0 as the baseline log-odds and $\phi_j^{(k)}$ representing the contribution of feature j in the k-th observation. Since SHAP values accurately capture each feature's influence for each observation, performing a logistic regression of these values on the binary target yields coefficients close to unity for those important features. In contrast, irrelevant features that do not consistently contribute across observations will have insignificant coefficients.

Lundberg and Lee (2017) transform the outputs of non-linear models into a linear format using SHAP, enabling straightforward feature ranking and visualization. However, in computer science, such ranked features or importance plots are often subject to manual selection. To address this limitation, we first train a machine learning model using all features in our dataset and extract SHAP values for each feature across more than 33,000 out-of-sample observations. Leveraging SHAP's additivity property, we then apply LASSO logistic regression to these SHAP values, incorporating an ℓ_1-penalty to shrink insignificant coefficients toward zero. The penalty parameter is determined through cross-validation, eliminating the need for manual intervention. By filtering out noise from irrelevant variables, the SHAP-LASSO approach enhances out-of-sample predictive performance. As a result, the SHAP-LASSO procedure selects a refined subset of predictors that not only improves model interpretability but also boosts predictive performance.

4.2 GS-LASSO Model

XGBoost (eXtreme Gradient Boosting) is a highly efficient implementation of the gradient boosting framework, which iteratively builds additive models by fitting decision trees to the residuals of previous trees. Its design makes it particularly well-suited for financial data, where complex interactions and non-linear relationships are common. Bogousslavsky et al. (2024) employ an XGBoost model to capture the nonlinear interactions among informed trading, volume, and volatility, proposing a novel measure derived from the model to quantify informed trading intensity. Similarly, Bartlett and O'Hara (2024) utilize an XGBoost model to detect the presence of undisplayed orders and optimize stock trade execution by predicting venues that offer superior price improvement. In this paper, we build our GS-LASSO model to classify the true direction of option trades (buy or sell) with high accuracy by applying XGBoost model using SHAP-LASSO for feature selection. This approach ensures robust and interpretable decision-making when distinguishing trade directions within complex market microstructure data.

4.2.1 GS-LASSO Model Construction

We initially construct 50 variables (listed in Table A1) that capture various sources of information related to option trades. These variables are categorized into three groups. The first group includes 21 option trade-specific features such as trade size, trade price, option moneyness, the midpoint rule effective spread and etc. The second group captures time-series information around the option trade execution, including volatility and returns before and after each trade at different time intervals (100 milliseconds, 1 second, and 5 seconds). The final group reflects stock market information, including the contemporaneous price of the VXX ETF, which is based on the VIX index and also time-series information around the stock trade execution, including volatility and returns before and after each trade at different time intervals (100 milliseconds, 1 second, and 5 seconds).

In feature selection, LASSO reduces the original set of 50 variables to 36 key predictors.

Among these, the dollar quoted spread, midpoint rule direction, and midpoint rule effective spread emerge as particularly influential, as indicated by the maximum absolute SHAP values commonly used in computer science.

To assess the robustness of our SHAP-LASSO and evaluate the relative importance of various features throughout the year, we conducted a 50-week SHAP–LASSO analysis of the whole year. The results, shown in Figure 4, indicate that most option trade–specific features were consistently selected across the entire year. Moreover, time–series information related to option trade execution was generally selected more frequently than that pertaining to stock trade execution.

Table A1 presents the features objectively selected by LASSO, along with the SHAP values for these 50 variables, ranked by their maximum absolute SHAP values. The top five features, based on their maximum absolute SHAP values, are: dollar quoted spread (2.66), trade size (2.65), option return during $[t-1s,t]$ (2.15), midpoint rule direction (1.86), and option price volatility during $[t-100ms,t]$ (1.8), . Meanwhile, the top five features, based on the mean SHAP values, are Moneyness (0.0109), Percentage Quoted Spread (0.0077), Effective Spread (ES) (0.0072), Dollar Quoted Spread (0.0068), and TQDiff (0.0039). Notably, Dollar Quoted Spread consistently appears among the top contributors across both metrics, underscoring its robust influence on the classification outcome.

It can be observed that LASSO selects the majority of features highly ranked by their absolute SHAP values. However, while univariate SHAP measures provide valuable insights, feature selection based solely on SHAP values may overlook certain contributing variables. In contrast, the LASSO procedure selects features based on their collective contribution to the model's explanatory power. For instance, although a univariate SHAP ranking might overlook the security type indicator, our LASSO-based approach successfully identifies its importance, highlighting its role in distinguishing between stock options and ETF options.

Our GS-LASSO Model also identifies market-wide volatility, measured by the VIX, as a significant variable. This feature might be overlooked if selection is based solely on the absolute

maximum or mean SHAP values during manual feature selection. Although VIX may not rank among the top features by SHAP value, its inclusion is justified by the fundamental relationship between option trading and overall market volatility. Furthermore, we observe that during our sample period, the accuracy of the midpoint rule is negatively correlated with the VIX; as stock market volatility increases, the midpoint rule's accuracy tends to decline. This suggests that during periods of heightened volatility, wholesalers may offer greater price improvements to manage and reduce their inventory risk.

4.2.2 Models Comparison Under Different Settings

We train three groups of XGBoost models and select the best-performing model based on out-of-sample accuracy. The first group consists of 21 option-related features from SLAN. The second group incorporates all 50 features, including time-series information surrounding option trades and market volatility proxies. The third group includes GS-LASSO Models utilizing a refined subset of 36 features.

For each group of models, we train using three different sample sizes: one day, one week (five trading days), and one month (22 trading days). Additionally, we compare two retraining strategies. In the static approach, the model is trained on a fixed window and then applied to subsequent out-of-sample data without updates. In contrast, the rolling approach continuously updates the training window by removing the oldest data point and incorporating the most recent observations. As a result, we train six XGBoost models for each group, yielding a total of 18 models across the three groups.

Theoretically, models utilizing SHAP-LASSO selected featuresshould outperform those using the full feature set, and rolling models are expected to deliver superior performance compared to static models, as they better capture evolving market dynamics. Consistent with this expectation, our one-week rolling GS-LASSO Model achieves the highest accuracy, with an average of 75.94% over the out-of-sample period from 9 January to 31 December.

Table 5 compares the performance of six GS-LASSO Models with different settings. These models vary by training-sample size (one day, one week, or one month) and by whether the data are updated in a rolling fashion. The one-week rolling model achieves the highest average accuracy among these models. specifically, as shown in the first row, the one-week rolling model exhibits significantly positive differences compared to each of the other five models. Moreover, among models that use SHAP-LASSO selected features, those employing a rolling retraining strategy consistently outperform their static counterparts, which aligns with our expectations.

Further analysis in Table 6 provides a detailed comparison between our best model, the one-week rolling GS-LASSO Model, and a naive midpoint-rule classifier. In this analysis, we partition the data into equal subsamples based on the median values of features such as price, size, dollar quoted spread, and midpoint effective spread, and then compare model performance across these subsamples. The results consistently show that our model outperforms the midpoint rule by approximately 20 percentage points.

Notably, in the 'price' subsamples, the below-median group shows that the midpoint rule achieves an accuracy of only 52.76%, indicating it is particularly unreliable for low-priced options compared to 60.21% accuracy in the above-median group. The low accuracy of the midpoint rule suggests that wholesalers tend to offer greater price improvements for these low-priced options trades. In contrast, our model improves classification accuracy for the below-median group by 27.83%, effectively leveraging trade price information to correct misclassifications in these low-priced options trades. Similarly, our model increases the accuracy by 27.37% in below-median 'dollar quoted spread' subsample from 54.89% to 82.26%. It is worth noting that while the GS-LASSO model enhances overall classification consistency, the improvements for below-median subsamples differ from those for above-median subsamples. This suggests that employing multiple GS-LASSO model sub-models tailored to different subsamples may further enhance accuracy.

5 When Wholesalers Provide Large Price Improvements?

Our SHAP-LASSO model significantly improves the accuracy of trade direction detection, enabling us to identify when wholesalers provide large PIs. In this section, we use PHLX true direction data to address two questions: do wholesalers profit after offering high PIs, and which types of SLAN trades receive higher PIs from wholesalers?

Figure 5 shows the empirical cumulative distribution function of price improvement based on 14.45 million SLAN trades. The data reveal that 17.84% of SLAN trades receive no PI, while 11.80% achieve full PIs, meaning investors buy at bid price and sell at ask price. Moreover, 7.56% of SLAN trades trade at midpoint, receiving a 50% PI, thus we observe a pronounced jump around 50%. These findings indicate that wholesalers have a strong incentive to offer significant PIs for certain SLAN trades. Moreover, wholesalers apply significantly different price improvement algorithms for stock and ETF options. For stock options, 15.38% of trades receive no PI, while 9.41% receive full improvement. In contrast, for ETF options, 28.53% of trades receive no PI and 22.18% receive full improvement. Additionally, wholesalers provide a price improvement of at least 50% for 30.77% of total trading volume, a level that implies losses on the bid-ask spread. This raises several important questions: Do wholesalers remain profitable after offering such high price improvements, and why do they offer large price improvements for certain trades?

To answer this, we assume that wholesalers close or hedge their option positions at the subsequent quote midpoint after a fixed time interval. For each SLAN trade i, we compute wholesalers dollar return as

$$\begin{aligned} Ret_i &= -100\text{Dir}_t \cdot \text{Volume}_t (P_{\text{midpoint},t+\tau} - P_{\text{actual},t}) \\ &= 100\text{Dir}_t \cdot \text{Volume}_t (P_{\text{actual},t} - P_{\text{midpoint},t+\tau}) \end{aligned} \tag{4}$$

where 100 is the multiplier for option contract, Dir_t is the true investor direction obtained from PHLX dataset. Volume_t is the number of traded contracts. $P_{\text{actual},t}$ is the SLAN trade price after

investor receive PI. $P_{\text{midpoint},t+\tau}$ is the corresponding options quote after τ seconds. In our analysis, we select τ as the next quote, 1 second, 5 seconds, 15 seconds, 30 seconds, 60 seconds, 90 seconds, and 120 seconds.

Table 7 compares wholesalers' profits under the assumption that they close or hedge their inventory at the quote midpoint at various time intervals (τ) after the trade. If wholesalers close positions at next updated quote midpoint after each SLAN execution, they earn an average daily profit of \$1.29 million, which is more than two times the \$0.58 million they earn if they close after 120 seconds. This monotonic decline in profit aligns with findings in the stock market (Conrad and Wahal (2020)). Across the entire sample, around 50% of trades are profitable even wholesalers close after 120 seconds, though this figure declines from 56.51% to 50.98% from next quote to 120 seconds. Weighting by volume shows that more than 70% of total volume proves profitable, suggesting that wholesalers selectively capture larger trades with favorable outcomes. The consistent gap between the count-based and volume-weighted profit fractions indicates that wholesalers exercise discretion in identifying lucrative opportunities: even if only slightly above half of all trades are winners, large-size winning trades dominate total volume and secure net gains.

We further decompose the wholesalers dollar return in Equation (4) into two parts: spread profit and position profit as follow

$$\begin{aligned} Ret_i &= 100\text{Dir}_t \cdot \text{Volume}_t (P_{\text{actual},t} - P_{\text{midpoint},t+\tau}) \\ &= \underbrace{100\text{Dir}_t \text{Volume}_t (P_{\text{actual},t} - P_{\text{midpoint},t})}_{\text{Spread Profit}} + \underbrace{100\text{Dir}_t \text{Volume}_t (P_{\text{midpoint},t} - P_{\text{midpoint},t+\tau})}_{\text{Position Profit}} \end{aligned} \tag{5}$$

Table 8 shows the components of wholesalers' profit based on the decomposition in equation (5). First, the profit is mainly driven by spread profit, which is determined at execution and remains unaffected by later option price movements. The daily average spread profit is \$1.33 million. Second, the overall position profit is negative across all time intervals, declining steadily from -\$0.36 million immediately after a quote update to -\$0.75 million after 120 seconds. Although

this negative profit is modest compared to the spread profit, it suggests that wholesalers could improve their results by closing or hedging positions more quickly. Additionally, the negative position profit may indicate that many SLAN traders are informed. Third, when splitting trades by price improvement (PI), below 50% versus 50% or above, we observe that trades with at least 50% PI generate positive position profit, while those with less than 50% PI yield negative profit. This pattern implies that wholesalers offer less PI to informed traders, leading to negative position profit, and more PI to uninformed traders, which allows them to earn positive position profit on those trades.

Table 9 presents the regression results exploring which types of SLAN trades receive higher price improvements (PIs) from wholesalers. The dependent variable is the percentage price improvement, as defined in equation (1), and the model includes fixed effects for both security and date. The buy indicator coefficient of 0.5238 suggests that wholesalers provide an extra 0.52% price improvement for buy trades. The significant call-put indicator coefficient of -0.3601 implies that call options receive 0.36% lower price improvements compared to put options. While the dollar quoted spread has a significant coefficient of 0.8654, a one standard deviation increase in the spread (0.59) results in only a 0.51% increase in price improvement. In summary, wholesalers tend to offer higher price improvements for buy put trades and wider dollar quoted spreads.

6 Conclusion

Due to large price improvements are now a commonplace feature of the options market, traditional stock-based classification rules, such as the midpoint rule, perform poorly in identifying correctly the direction of option trades. To overcome this problem, we propose a GS-LASSO model, the machine learning model XGBoots combined with LASSO selection by the SHAP, which increases classification accuracy from approximately 56% to nearly 76%. This methodological advancement appears to have significant implications: it enhances the precision of empirical studies on trading costs, spreads, and execution quality, and it clarifies that the scale of price improvement can remain

high without eroding wholesaler profitability.

Our results further indicate that the U.S. options market, which has long operated under an order-by-order auction regime, provides an instructive analogue for the SEC's proposed reforms in the stock market. The evidence suggests that investors are likely to benefit from substantial price improvements when market makers compete for individual orders, in keeping with auction theoretic predictions of increased allocative efficiency.

In short, this paper not only develops an effective machine-learning method for accurately classifying retail option trades but also provides useful insights for regulators, highlighting the important role that competitive auctions play in delivering significant price improvements to retail investors.

References

Amihud, Y., Mendelson, H., 1980. Dealership market: Market-making with inventory. Journal of financial economics 8, 31–53.

Anand, A., Muravyev, D., 2024. Do auctions impact quote competition? Available at SSRN 4891227 .

Avramov, D., Cheng, S., Metzker, L., 2023. Machine learning vs. economic restrictions: Evidence from stock return predictability. Management Science 69, 2587–2619.

Bartlett, R. P., O'Hara, M., 2024. Navigating the murky world of hidden liquidity. Available at SSRN .

Bogousslavsky, V., Fos, V., Muravyev, D., 2024. Informed trading intensity. The Journal of Finance 79, 903–948.

Bryzgalova, S., Pavlova, A., Sikorskaya, T., 2023. Retail trading in options and the rise of the big three wholesalers. The Journal of Finance 78, 3465–3514.

Chen, L., Pelger, M., Zhu, J., 2024. Deep learning in asset pricing. Management Science 70, 714–750.

Chinco, A., Clark-Joseph, A. D., Ye, M., 2019. Sparse signals in the cross-section of returns. The Journal of Finance 74, 449–492.

Conrad, J., Wahal, S., 2020. The term structure of liquidity provision. Journal of Financial Economics 136, 239–259.

Ellis, K., Michaely, R., O'Hara, M., 2000. The accuracy of trade classification rules: Evidence from nasdaq. Journal of financial and Quantitative Analysis 35, 529–551.

Ernst, T., Spatt, C. S., 2024. Payment for order flow and option internalization. Available at SSRN 4056512 .

Ernst, T., Spatt, C. S., Sun, J., 2022. Would order-by-order auctions be competitive? Available at SSRN 4300505 .

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. The Review of Financial Studies 33, 2326–2377.

Garman, M. B., 1976. Market microstructure. Journal of financial Economics 3, 257–275.

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

Han, Y., He, A., Rapach, D. E., Zhou, G., 2024. Cross-sectional expected returns: New fama–macbeth regressions in the era of machine learning. Review of Finance 28, 1807–1831.

Hendershott, T., Khan, S., Riordan, R., 2022. Option auctions. Available at SSRN 4110516 .

Ho, T., Stoll, H. R., 1981. Optimal dealer pricing under transactions and return uncertainty. Journal of Financial economics 9, 47–73.

Holden, C. W., Pierson, M., Wu, J., 2023. In the blink of an eye: Exchange-to-sip latency and trade classification accuracy. Available at SSRN 4441422 .

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271–292.

Lee, C. M., Ready, M. J., 1991. Inferring trade direction from intraday data. The Journal of Finance 46, 733–746.

Light, N., Maslov, D., Rytchkov, O., 2017. Aggregation of information about the cross section of stock returns: A latent variable approach. The Review of Financial Studies 30, 1339–1381.

Lundberg, S. M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874 .

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

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

Savickas, R., Wilson, A. J., 2003. On inferring the direction of option trades. Journal of Financial and Quantitative Analysis 38, 881–902.

Table 1 Descriptive Statistics of Daily SLAN Trades

This table presents the descriptive statistics of daily SLAN trades. Trade Amount and Volume are reported in millions. ATM, ITM, and OTM volumes correspond to options classified by moneyness. "Stock Option" and "ETF Option" represent stock and ETF options volume. The Trade Size is calculated as the average of the daily average trade sizes across all SLAN trades. Days to Expiration is computed by summing each trade's expiration time (in years) weighted by its trade size and then multiplying by 365. The percentage quoted spread is daily volume-weighted average percentage quoted spread. The call trade percentage and buy trade percentage measure the proportions of call option trades and buy-side trades, respectively. Data period covers from January 1, 2024 to December 31, 2024.

	Mean	Std	Min	25%	Median	75%	Max
Trade Amount (Thousand)	185.24	37.21	57.91	162.71	181.05	204.98	317.34
Volume (Million)	1.57	0.30	0.50	1.37	1.55	1.76	2.38
ATM Volume (Million)	0.96	0.21	0.29	0.82	0.95	1.11	1.62
ITM Volume (Million)	0.22	0.05	0.08	0.19	0.21	0.25	0.42
OTM Volume (Million)	0.38	0.09	0.13	0.32	0.37	0.43	0.73
Stock Option (Million)	1.14	0.24	0.37	0.96	1.11	1.28	1.80
ETF Option (Million)	0.43	0.09	0.13	0.37	0.43	0.49	0.68
Trade Size (Contracts/Trade)	8.49	0.64	6.82	8.09	8.46	8.87	10.37
Days to Expiration	31.54	3.91	22.73	28.94	31.34	33.87	48.74
Percentage Quoted Spread	0.1509	0.0260	0.1022	0.1314	0.1481	0.1677	0.2470
Call Trade Percentage	65.00%	2.95%	58.33%	62.78%	64.95%	67.33%	71.86%
Buy Trade Percentage	50.61%	1.71%	44.72%	49.87%	50.83%	51.71%	54.06%

Table 2 Descriptive Statistics of SLAN Price Improvements

This table presents the descriptive statistics of trade-level SLAN trade price improvements. The price improvement is calculated using the true trading direction (TTD) obtained from PHLX. The percentage price improvement follows Equation (1), while the dollar price improvement corresponds to the numerator in (1). The statistics are based on SLAN trades executed within the quoted spread. Data period covers from January 1, 2024 to December 31, 2024.

	N	Mean	Std	5%	25%	Median	75%	95%
Panel A: Percentage Price Improvement								
All Trades	46.68M	43.2%	30.3%	0.0%	20.0%	40.0%	60.0%	100.0%
Buy Trades	23.22M	43.5%	30.6%	0.0%	20.0%	42.2%	60.0%	100.0%
Sell Trades	23.47M	43.0%	30.0%	0.0%	20.8%	40.0%	60.0%	100.0%
Call Trades	30.54M	43.1%	29.9%	0.0%	20.0%	40.0%	60.0%	100.0%
Put Trades	16.14M	43.5%	31.0%	0.0%	20.0%	42.1%	60.0%	100.0%
Panel B: Dollar Price Improvement (Unit: $)								
All Trades	46.68M	0.087	0.295	0	0.01	0.02	0.06	0.32
Buy Trades	23.22M	0.086	0.297	0	0.01	0.02	0.06	0.32
Sell Trades	23.47M	0.088	0.293	0	0.01	0.02	0.06	0.33
Call Trades	30.54M	0.091	0.304	0	0.01	0.02	0.07	0.34
Put Trades	16.14M	0.079	0.277	0	0.01	0.02	0.06	0.30

Table 3 Intraday Accuracy Regression across securities and time

This table reports the intraday accuracy regression

$$\text{Accuracy}_{t,i} = \beta_0 + \beta_1 \text{Spread}_{t,i} + \beta_2 log(\text{Volume})_{t,i} + \beta_3 \text{VXX}_{t,i} + \text{Fixed Effects} + \varepsilon_{t,i},$$

where $\text{Accuracy}_{t,i}$ is the midpoint rule accuracy for the ith half-hour interval on day t. $\text{Spread}_{t,i}$ is volume-weighted average dollar quoted spread for the corresponding half-hour interval. $log(\text{Volume})_{t,i}$ is the log of the volume. $\text{VXX}_{t,i}$ is the VIX ETF midpoint quote at the end of the ith half-hour interval on day t. The standard deviations are clustered by time interval. Significance is denoted by ***, **, and * for the 1%, 5%, and 10% levels, respectively. The dataset includes all trades executed between January 1, 2024, and December 31, 2024.

Variable	All	Stock Options	ETF Options
$\text{Spread}_{t,i}$	0.0543	-0.0086	0.2420***
	(1.57)	(-0.36)	(3.21)
$log(\text{Volume})_{t,i}$	0.0161***	0.0156***	0.0199***
	(4.98)	(4.55)	(13.15)
$\text{VXX}_{t,i}$	0.0002***	-0.0001**	0.0012***
	(5.66)	(-2.37)	(17.52)
Interval Fixed Effects	Yes	Yes	Yes
Observations	3259	3256	3257
Adj R^2	0.1619	0.1827	0.2253

Table 4 Average Accuracy of Midpoint Rule Matching with Various Quotes

This table presents the average accuracy of midpoint rule matching using quotes from various time points: 1 second before and after the trade, and 100 milliseconds before the trade. The 'Contemporary Quote' refers to the quote provided in the OPRA data, while the last quote before the trade and next quote after the trade represent the most recent NBBO updates prior to and following the trade. Data period covers from January 1, 2024 to December 31, 2024.

Time point	Mean	Std	Min	Max
Panel A: Stock options and ETF options				
1s before trade	51.52%	2.15%	42.89%	56.25%
100ms before trade	55.95%	2.45%	45.80%	61.07%
Last quote before trade	56.31%	2.41%	46.58%	61.31%
contemporary	56.51%	2.39%	46.90%	61.54%
Next quote after trade	56.47%	2.38%	46.90%	61.43%
1s after trade	55.32%	2.04%	47.55%	60.28%
Panel B: Stock options				
1s before trade	52.20%	2.04%	45.63%	56.90%
100ms before trade	56.72%	2.19%	49.84%	61.91%
Last quote before trade	56.91%	2.17%	50.11%	62.13%
contemporary	57.10%	2.14%	50.38%	62.23%
Next quote after trade	57.02%	2.13%	50.32%	62.19%
1s after trade	55.90%	1.90%	49.96%	61.10%
Panel C: ETF options				
1s before trade	48.39%	3.70%	35.15%	55.91%
100ms before trade	52.39%	5.36%	36.41%	66.89%
Last quote before trade	53.43%	5.31%	37.89%	67.36%
contemporary	53.70%	5.30%	38.31%	67.49%
Next quote after trade	53.83%	5.28%	38.48%	67.51%
1s after trade	52.61%	3.93%	40.77%	60.18%

Table 5 Out-of-Sample Performance Across Various SHAP-LASSO Models

This table reports the out-of-sample performance for all pairwise model comparisons. The models include 1M Static, 1M Rolling, 1W Static, 1W Rolling, 1D Static, and 1D Rolling, which differ in their training approach. "Static" uses a fixed training period, while "Rolling" is updated periodically. Each entry shows the daily average accuracy difference between the row model and the column model with the corresponding t-statistic in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively. Each cell in the table reports the difference of the row model minus that of the column model. Data period covers from January 1, 2024 to December 31, 2024.

	1M Rolling	1D Rolling	1M Static	1W Static	1D Static
1W Rolling	0.21%	0.25%	0.61%***	0.65%***	0.73%***
	(1.21)	(1.38)	(3.55)	(3.58)	(3.97)
1M Rolling		0.04%	0.41%**	0.44%**	0.52%***
		(0.22)	(2.36)	(2.45)	(2.86)
1D Rolling			0.37%**	0.40%**	0.48%**
			(2.04)	(2.14)	(2.54)
1M Static				0.03%	0.12%
				(0.19)	(0.64)
1W Static					0.08%
					(0.43)

Table 6 Model Accuracy Comparison Across Various Feature Subsamples

Table presents the various subsamples accuracy of two models: the GS-LASSO 1W Rolling model and the midpoint model. The results are split sample into two subsamples based on the median of each feature, labeled "Below Median" and "Above Median". For each feature, the table reports the accuracy for both models, the difference (calculated as 1W Rolling minus midpoint), and the corresponding t-statistic (shown in parentheses). Significance is denoted by ***, **, and * for the 1%, 5%, and 10% levels, respectively. Data period covers from January 1, 2024 to December 31, 2024.

	Below Median			**Above Median**		
Features	**GS-LASSO**	**midpoint**	**Difference**	**GS-LASSO**	**midpoint**	**Difference**
	(1)	(2)	(1)-(2)	(3)	(4)	(3)-(4)
Size	75.63%	57.57%	18.07%***	76.40%	54.91%	21.48%***
			(135.10)			(125.71)
Price	80.59%	52.76%	27.83%***	71.28%	60.21%	11.07%***
			(106.69)			(151.93)
YExp	78.56%	55.75%	22.81%***	73.33%	57.21%	16.12%***
			(115.00)			(146.75)
Moneyness	75.92%	56.35%	19.57%***	75.96%	56.60%	19.36%***
			(125.42)			(129.90)
Dollar Quoted Spread	82.26%	54.89%	27.37%***	69.31%	58.16%	11.15%***
			(92.70)			(116.46)
Bid-Ask Depth Ratio	75.89%	56.38%	19.51%***	75.99%	56.57%	19.42%***
			(134.13)			(127.50)
Midpoint ES	71.29%	49.53%	21.76%***	80.61%	63.44%	17.17%***
			(216.97)			(80.35)
Quote Trade Time Diff	78.28%	57.96%	20.32%***	73.59%	54.98%	18.61%***
			(122.91)			(137.42)

Table 7 Wholesalers Daily Estimated Profits from SLAN Trades

This table reports the daily average profit of wholesalers, calculated based on executed trades and the prevailing midpoint price at different time intervals after the trade. Daily Avg Profits refers to the average aggregated profit earned by wholesalers each day over the sample period. Profit per trade represents the profit on each trade, Profit Trade Percentage denotes the fraction of trades yielding profits, and Profit Volume Percentage represents the total size of profitable trades relative to all trades where the post-trade midpoint is observable. The dataset includes all trades executed between January 1, 2024 to December 31, 2024.

Midpoint after trade	Daily Avg Profit	Profit/Trade	Profit Trade Pct	Profit Volume Pct
Next quote updated	$1,290,105	$6.96	56.51%	70.81%
1 second after trade	$1,189,706	$6.42	55.37%	68.28%
5 second after trade	$1,063,515	$5.74	53.74%	63.20%
15 second after trade	$900,621	$4.86	52.58%	58.34%
30 second after trade	$768,084	$4.15	51.92%	55.64%
60 second after trade	$672,472	$3.63	51.38%	53.77%
90 second after trade	$662,820	$3.58	51.13%	53.05%
120 second after trade	$580,611	$3.13	50.98%	52.52%

Table 8 Decomposition of Wholesalers Profit from SLAN Trades

This table reports the components of daily wholesalers profit. The dataset includes all trades executed between January 1, 2024 to December 31, 2024.

			Position Profit		
Midpoint after trade	Daily Ave Profit	Spread Profit	Overall	PI$\geq$ 50%	PI$<$ 50%
Next quote updated	\$1,290,105	\$1,326,187	-\$36,082	\$36,271	-\$72,353
1 second after trade	\$1,189,706	\$1,326,187	-\$136,481	\$141,118	-\$277,599
5 second after trade	\$1,063,515	\$1,326,187	-\$262,673	\$244,987	-\$507,659
15 second after trade	\$900,621	\$1,326,187	-\$425,567	\$327,389	-\$752,956
30 second after trade	\$768,084	\$1,326,187	-\$558,103	\$374,762	-\$932,865
60 second after trade	\$672,472	\$1,326,187	-\$653,715	\$430,640	-\$1,084,355
90 second after trade	\$662,820	\$1,326,187	-\$663,367	\$498,163	-\$1,161,530
120 second after trade	\$580,611	\$1,326,187	-\$745,577	\$452,820	-\$1,198,397

Table 9 Price Improvement Regression with Securities and Day Fixed Effects

This table employs a fixed effects regression model to control for unobserved heterogeneity across both securities (indexed by i) and time (indexed by t). The general specification of the model is given by:

$$\begin{aligned} \text{PI}_i = \beta_0 + \beta_1 \text{cp}_i + \beta_2 \text{size}_i + \beta_3 \text{price}_i \\ + \beta_4 \text{buyind}_i + \beta_5 \text{yexp}_i + \beta_6 \text{bs_spread}_i + \beta_7 \text{tqdiff}_i + \varepsilon_t, \end{aligned} \tag{6}$$

where PI_i is the percentage price improvement of trade i, measured as united of percentage. cp_i is an indicator variable which takes 1 when the trade is call, 0 otherwise; size_i is the trade size; price_i is the trade price; buyind_i is a buy indicator which takes 1 when it is buy trade, 0 otherwise; yexp_i is time to expiration; bs_spread_i is the dollar quoted spread; tqdiff_i represents quote trade time difference defined in Table A1. The standard deviations are clustered by date. Significance is denoted by ***, **, and * for the 1%, 5%, and 10% levels, respectively. The dataset includes all trades executed between January 9, 2024, and December 31, 2024.

Variable	(1)	(2)	(3)
cp	-0.3601***	-0.5231***	-0.3725***
	(-5.34)	(-6.26)	(-5.50)
size	-0.0400***	-0.0385***	-0.0401***
	(-9.16)	(-9.18)	(-9.15)
price	0.0178***	0.0111***	0.0195***
	(10.45)	(6.99)	(11.73)
buyind	0.5238***	0.5338***	0.5058***
	(10.64)	(9.88)	(10.07)
yexp	0.6134***	0.8273***	0.5988***
	(7.51)	(7.77)	(7.16)
bs spread	0.8654***	1.0193***	0.8393***
	(21.99)	(17.36)	(23.17)
tqdiff	-0.0004***	-0.0007***	-0.0004***
	(-15.00)	(-21.16)	(-15.14)
Security Fixed Effects	Yes	No	Yes
Day Fixed Effects	Yes	Yes	No
Observations	46.68M	46.68M	46.68M
Adj R^2	0.0195	0.0083	0.0161

Figure 1. Time Series of Midpoint Rule Accuracy Rate

The figure presents the time series of accuracy rate for the midpoint rule, volatility midpoint rule and our SHAP-LASSO XGBoost model. Data period covers from January 1, 2024 to December 31, 2024.

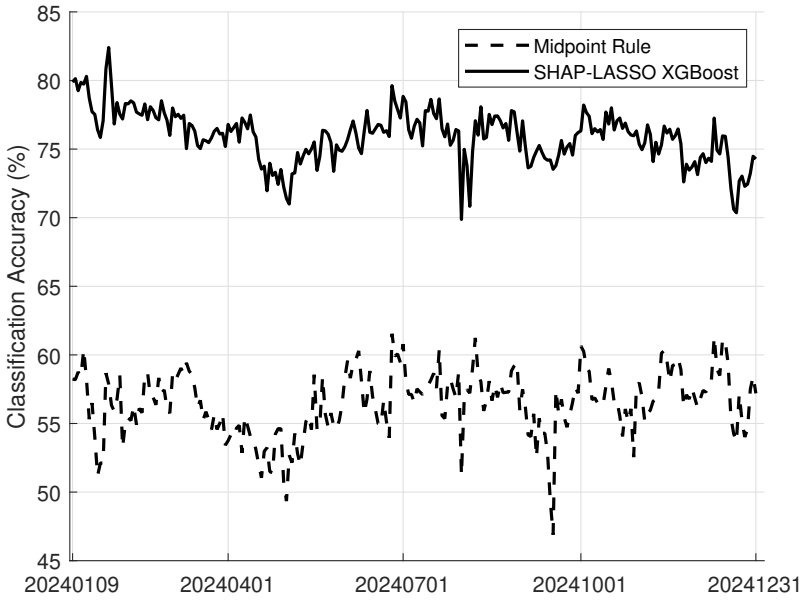


Figure 2. Intraday Accuracy Rate of Midpoint Rule

The figures present the intraday patterns of midpoint rule. Data period covers from January 1, 2024 to December 31, 2024.

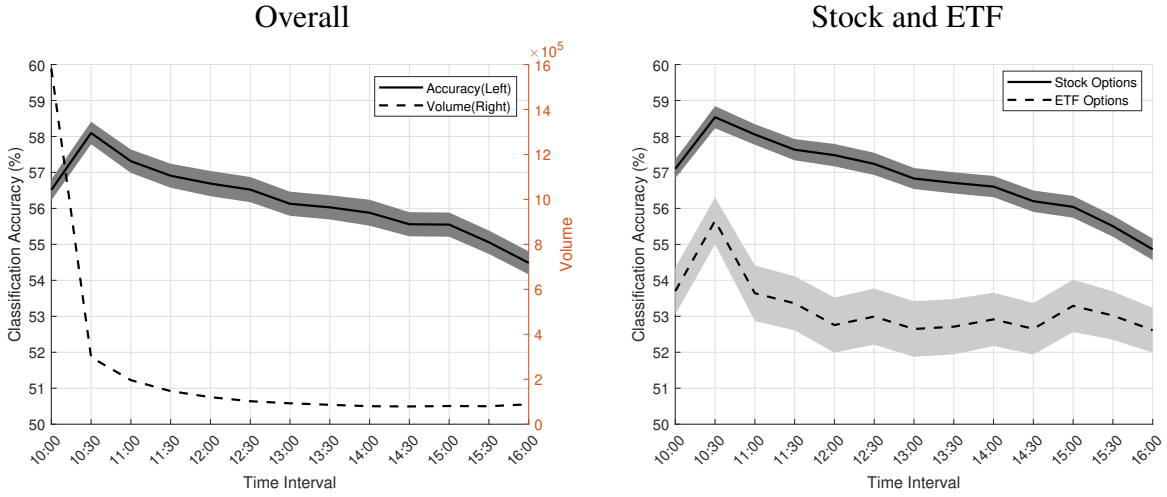


Figure 3. Duration of PHLX Auction in Data

The figure presents the time series of the duration between two timestamps: PHLX's auction message and the execution record on OPRA. Data period covers from January 1, 2024 to December 31, 2024.

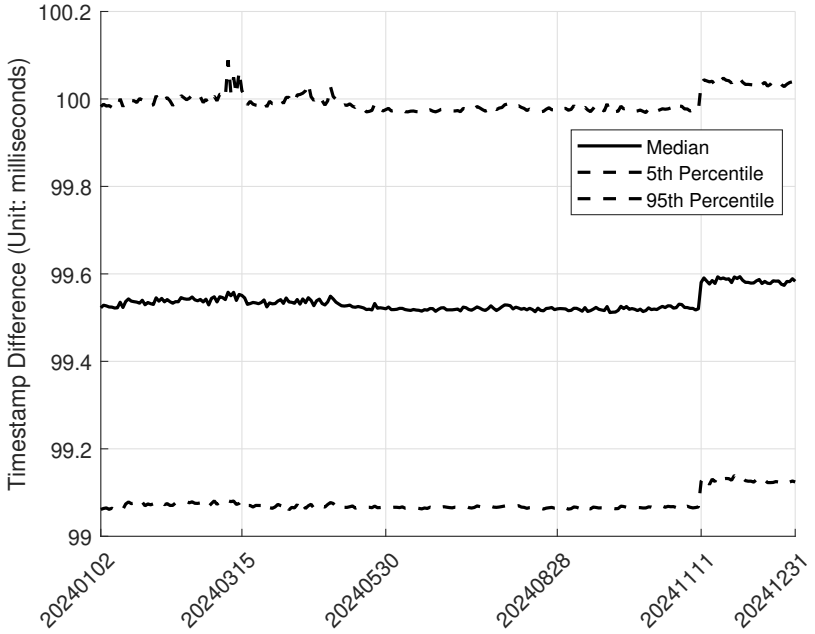


Figure 4. Variable Selection from SHAP-LASSO

The figure displays the weekly variable selection from SHAP-LASSO. Data period covers from January 1, 2024 to December 31, 2024.

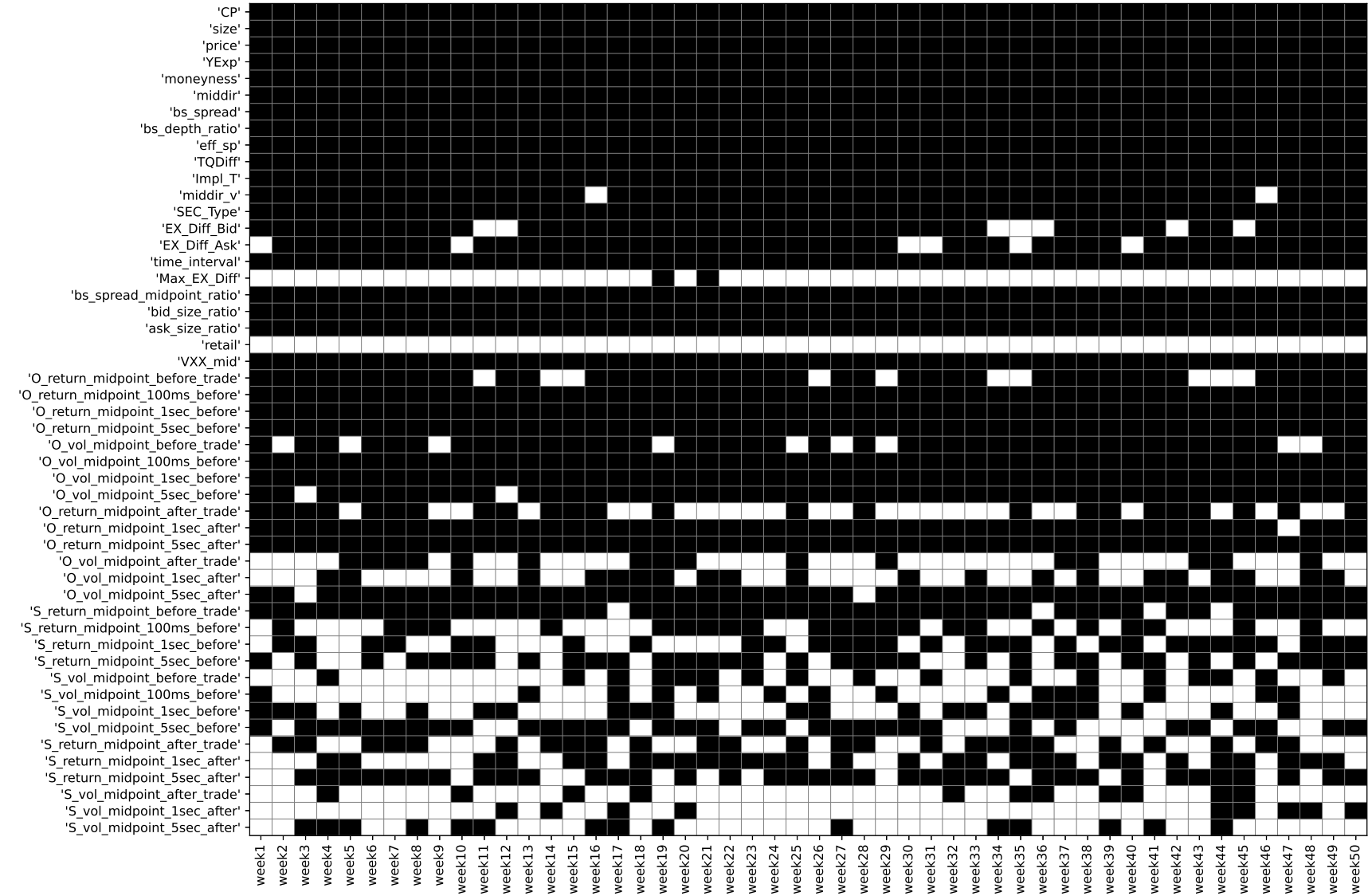
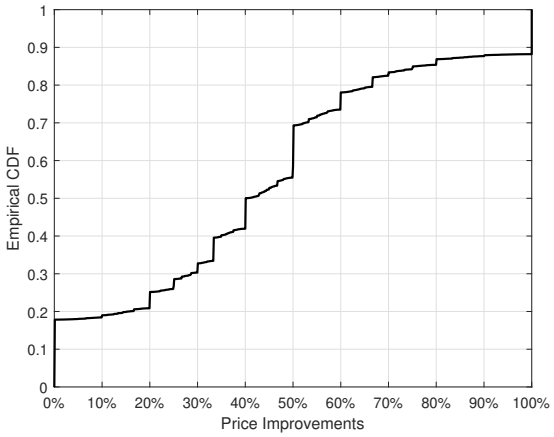


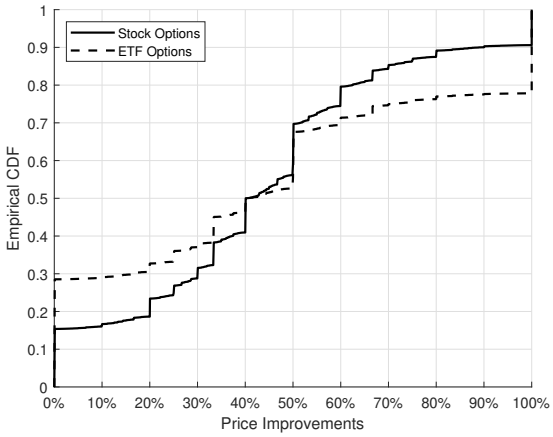
Figure 5. Empirical Cumulative Distribution Function of Price Improvements

The figures display the empirical cumulative distribution function of price improvements obtained from SLAN investors. Price improvements are computed using equation (1). Data period covers from January 1, 2024 to December 31, 2024.

Overall



Stock and ETF



Internet Appendix

1 Variables List Used in GS-LASSSO Model

In this secton, we present each feature's SHAP values using *All Features* in the model and features selected by LASSO. In the Feature Name column, *Ret* represents return while *vol* represents volatility. We report the `Absolute Max` SHAP value (rounded to two decimals) and the `Mean` SHAP value (rounded to four decimals; very small values are shown in scientific notation). A checkmark in the SHAP-LASSO column indicates that the feature was selected by LASSO, while a dash ("–") indicates that it was not included. Data period covers from January 1, 2024 to January 5, 2024.(Week1)

Table A1 SHAP Values for All Features Model

Feature Name	Description	SHAP-LASSO	Absolute Max	Mean
Dollar Quoted Spread	$P_{\text{Ask},t} - P_{\text{Bid},t}$	✓	2.66	0.0068
Size	number of contracts	✓	2.65	-0.0021
$O_Ret_{[t-1s,t]}$	$\log(O_MidP_t) - \log(O_MidP_{t-1s})$	✓	2.15	-0.0086
Midpoint Rule Direction	Detect trade direction based on midpoint	✓	1.86	-0.0058
$O_Ret_{[t-100ms,t]}$	$\log(O_MidP_t) - \log(O_MidP_{t-100ms})$	✓	1.80	-0.0074

Continued on next page

Table A1 – continued from previous page

Feature Name	Description	SHAP-LASSO	Absolute Max	Mean
$O_Vol_{[t-100ms,t]}$	$\left\lvert \log(O_MidP_t) - \log(O_MidP_{t-100ms}) \right\rvert$	✓	1.73	-0.0019
Percentage Quoted Spread	$\frac{P_{\text{Ask},t} - P_{\text{Bid},t}}{\text{MidP}_t}$	✓	1.72	0.0077
Effective Spread(ES)	$\frac{\left\lvert P_{\text{trade}} - \text{MidP} \right\rvert}{\text{MidP}}$	✓	1.24	0.0072
Years to expiry	Time to expiry (in years)	✓	1.11	-8.0893×10^{-4}
Bid-Ask Depth Ratio	$\frac{Q_{\text{Bid}}}{Q_{\text{Ask}}}$	✓	1.10	0.0010
Price	Trade Price	✓	0.82	-0.0053
TQDiff	Difference between the last quote timestamp and the trade timestamp	✓	0.75	0.0039
Moneyness	$\frac{\text{Strike Price}}{\text{Underlying Price}}$	✓	0.68	0.0109
$O_Ret_{[t,t+\Delta s]}$	$\log(O_MidP_{t+\Delta s}) - \log(O_MidP_t)$	✓	0.59	6.1549×10^{-4}
$O_Ret_{[t-\Delta s,t]}$	$\log(O_MidP_t) - \log(O_MidP_{t-\Delta s})$	✓	0.58	0.0021
$O_Vol_{[t-\Delta s,t]}$	$\left\lvert \log(O_MidP_t) - \log(O_MidP_{t-\Delta s}) \right\rvert$	✓	0.51	0.0012
C/P Indicator	equals 1 for call, 0 otherwise	✓	0.45	-0.0099
Volatility Midpoint Rule Direction	Detect trade direction based on volatility midpoint	✓	0.44	0.0021
Bid Size Ratio	$\frac{Q_{\text{Bid}}}{\text{Size}}$	✓	0.43	-3.3969×10^{-4}
$S_Ret_{[t-\Delta s,t]}$	$\log(S_MidP_t) - \log(S_MidP_{t-\Delta s})$	✓	0.41	1.0065×10^{-4}

Continued on next page

Table A1 – continued from previous page

Feature Name	Description	SHAP-LASSO	Absolute Max	Mean
Quote Trade Time	difference between quote timestamp and trade timestamp	✓	0.39	-0.0014
$O_Vol_{[t-1s,t]}$	$\left\lvert \log(O_MidP_t) - \log(O_MidP_{t-1s}) \right\rvert$	✓	0.39	6.8257×10^{-4}
$O_Ret_{[t-5s,t]}$	$\log(O_MidP_t) - \log(O_MidP_{t-5s})$	✓	0.38	-2.9525×10^{-4}
$O_Ret_{[t,t+5s]}$	$\log(O_MidP_{t+5s}) - \log(O_MidP_t)$	✓	0.32	1.2722×10^{-4}
$O_Vol_{[t,t+\Delta s]}$	$\left\lvert \log(O_MidP_{t+\Delta s}) - \log(O_MidP_t) \right\rvert$	–	0.28	3.1443×10^{-5}
Implied Volatility	trade implied volatility	✓	0.27	-0.0045
Ask size ratio	ask size / total size	✓	0.26	5.5086×10^{-4}
$O_Vol_{[t-5s,t]}$	$\left\lvert \log(O_MidP_t) - \log(O_MidP_{t-5s}) \right\rvert$	✓	0.24	0.0019
Security Type	1 for stock options, 2 for ETF options	✓	0.19	4.0973×10^{-4}
$S_Ret_{[t-5s,t]}$	$\log(S_MidP_t) - \log(S_MidP_{t-5s})$	✓	0.16	0.0017
$O_Ret_{[t,t+1s]}$	$\log(O_MidP_{t+1s}) - \log(O_MidP_t)$	✓	0.16	-5.5570×10^{-4}
$O_Vol_{[t,t+1s]}$	$\left\lvert \log(O_MidP_{t+1s}) - \log(O_MidP_t) \right\rvert$	–	0.15	5.1971×10^{-5}
$S_Ret_{[t,t+5s]}$	$\log(S_MidP_{t+5s}) - \log(S_MidP_t)$	–	0.15	0.0018
VXX Price	midpoint price for VXX ETF	✓	0.15	8.2664×10^{-4}
$O_Vol_{[t,t+5s]}$	$\left\lvert \log(O_MidP_{t+5s}) - \log(O_MidP_t) \right\rvert$	✓	0.13	-9.5926×10^{-4}
$S_Vol_{[t-5s,t]}$	$\left\lvert \log(S_MidP_t) - \log(S_MidP_{t-5s}) \right\rvert$	✓	0.11	3.6149×10^{-4}

Continued on next page

Table A1 – continued from previous page

Feature Name	Description	SHAP-LASSO	Absolute Max	Mean
Bid Exchange Difference	1 if NBB is not from PHLX	✓	0.07	1.9148×10^{-4}
$S_Vol_{[t-\Delta s,t]}$	$\left\lvert \log(S_MidP_t) - \log(S_MidP_{t-\Delta s}) \right\rvert$	–	0.06	0.0012
$S_Ret_{[t-1s,t]}$	$\log(S_MidP_t) - \log(S_MidP_{t-1s})$	–	0.05	0.0016
$S_Vol_{[t-1s,t]}$	$\left\lvert \log(S_MidP_t) - \log(S_MidP_{t-1s}) \right\rvert$	✓	0.05	0.0013
$S_Ret_{[t,t+1s]}$	$\log(S_MidP_{t+1s}) - \log(S_MidP_t)$	–	0.04	4.6719×10^{-4}
$S_Ret_{[t,t+\Delta s]}$	$\log(S_MidP_{t+\Delta s}) - \log(S_MidP_t)$	–	0.04	5.0076×10^{-4}
Ask Exchange Difference	1 if NBO is not from PHLX	–	0.03	9.3594×10^{-6}
$S_Ret_{[t-100ms,t]}$	$\log(S_MidP_t) - \log(S_MidP_{t-100ms})$	–	0.03	5.3144×10^{-4}
Quote Exchange Difference	0 if NBBO are from PHLX	–	0.02	2.8839×10^{-6}
$S_Vol_{[t-100ms,t]}$	$\left\lvert \log(S_MidP_t) - \log(S_MidP_{t-100ms}) \right\rvert$	✓	0.01	4.8412×10^{-5}
NYSE Retail Indicator	1 if size ≤ 10	–	0.00	0
$S_Vol_{[t,t+\Delta s]}$	$\left\lvert \log(S_MidP_{t+\Delta s}) - \log(S_MidP_t) \right\rvert$	–	0.00	0
$S_Vol_{[t,t+1s]}$	$\left\lvert \log(S_MidP_{t+1s}) - \log(S_MidP_t) \right\rvert$	–	0.00	0
$S_Vol_{[t,t+5s]}$	$\left\lvert \log(S_MidP_{t+5s}) - \log(S_MidP_t) \right\rvert$	–	0.00	0