Inferring Trade Directions in Options via Machine Learning

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

Current Version: August 2025

*Jindal School of Management, University of Texas at Dallas, si.gao@utdallas.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
§University at Albany - SUNY, Massry School of Business, zzhou22@albany.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 up to 75.94% accuracy, consistently 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 Extreme Gradient Boosting (XGBoost), SHapley Additive Explanations (SHAP), and Least Absolute Shrinkage and Selection Operator (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.[1] Our dataset covers

[1] In 2023, PHLX held an 11.29% market share in equity options, second only to ARCA's 12.18%. Source: OCC (https://www.theocc.com/market-data/market-data-reports/volume-and-open-interest/volume-by-exchange).

January 1, 2024, to December 31, 2024, and includes 49.72 million detailed auction messages, transaction-level data from Options Price Reporting Authority (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.[2] 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.[3] 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. The simplest

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

[3]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.

XGBoost model achieves an out-of-sample accuracy of 73.65%. To enhance both accuracy and interpretability, we apply SHAP (Lundberg and Lee (2017)) to explain the non-linear model and introduce a novel feature selection method using LASSO on SHAP values,and then using the selected features to build our GS-LASSO model. This GS-LASSO model further improves accuracy to up 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 features. 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.[4] 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

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

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 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

2.1 Data Description

Our study primarily utilizes auction messages extracted from the Philadelphia Stock Exchange (PHLX), which ranked second among the 17 U.S. options exchanges by market share for equity options in 2023. Specifically, we collect historical data from the PHLX order feed, which captures the status of simple and complex orders 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. For our analysis, we extract all simple order price improvement auction messages, including order direction (buy or sell), option contract specifications, timestamps, auction prices, and auction volumes. We subsequently match these

orders with transaction-level data from the Options Price Reporting Authority (OPRA), provided via ThetaData, using timestamps and option contract identifiers.

We then construct the Single-Leg Auction Non-ISO Trade (SLAN) dataset by identifying trades with the SLAN trade condition from the OPRA dataset, consistent with the retail options trade classification in Bryzgalova et al. (2023). Our dataset comprises 49.72 million SLAN trades executed on PHLX, along with their true trade direction. In 2024, SLAN trades accounted for 63.10% of the total options volume on PHLX, closely aligning with Bryzgalova et al. (2023), who estimated SLAN trades comprised over 60% of total market volume in early 2020. This suggests that our dataset is representative of overall market activity. To further enhance our analysis, we match the SLAN 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 summary statistics of daily trades for the SLAN dataset only contains stock option and ETF option data spanning from January 1, 2024, to December 31, 2024. To ensure data quality, we apply several filters: (1) removing trades executed outside the quoted spread (prices above ask or below bid); (2) excluding observations with ask prices less than or equal to bid prices; (3) discarding trades with non-positive prices; and (4) filtering out trades with non-positive sizes or sizes exceeding 250,000 contracts.[5] These exclusions remove only around 0.93% of the total observations. After retaining only stock and ETF option data and filtering the data, the dataset volume decreases from 49.72 million to 46.68 million SLAN trades. In statistics, the average daily trading volume is approximately 1.57 million contracts, ranging from 0.5 million to 2.38 million, with the peak volume more than four times the lowest recorded value, indicating substantial variation in trading activity across days.

We classify options as at-the-money (ATM), in-the-money (ITM), or out-of-the-money (OTM) based on quotes at the precise trade time, rather than end-of-day prices. ATM options, defined as those with moneyness between 0.95 and 1.05, represent the majority of the trading activity,

[5]The limit of 250,000 contracts is set by NASDAQ PHLX's position rules. See Section 13(g) at https://listingcenter.nasdaq.com/rulebook/phlx/rules/.

accounting for approximately 61.54% of total trading volume. OTM options constitute about 24.55%, while ITM options account for 13.88%. Call options dominate the dataset in both volume and number of trades, with approximately 65.00% of total volume and a similar proportion in trade count, consistent with the finding in CBOE's retail trading report in 2024.[6] Additionally, buy and sell volumes are nearly balanced, with buy orders accounting for about 50.61% of SLAN trades. Finally, stock and ETF options together represent 94.64% of total SLAN trades on PHLX; specifically, stock options account for 68.57%, and ETF options comprise 26.07%, highlighting the growing popularity of ETF options in recent years. The remaining 5.36% includes index options, ETPs, ADRs, and other instruments.

Interestingly, ETF options and stock options exhibit distinct trading patterns in our sample. First, SLAN trades for ETF options are significantly larger on average compared to stock options. Specifically, stock options account for approximately 150,553 trades per day, representing roughly 80% of total daily trades, whereas ETF options average around 34,691 trades, or about 20%. Consequently, the average trade size for ETF options is 12.68 contracts, notably exceeding the average of 7.57 contracts 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 aligns with NYSE's operational definition, where trades of no more than 10 contracts predominantly indicate retail participation, suggesting retail investors comprise the majority of SLAN activity.[7]

Second, the quoted spreads for ETF options are significantly tighter than those for stock options. The average percentage quoted spread in our sample is 0.1509, slightly lower than the 0.172 reported by Muravyev and Pearson (2020), who analyzed S&P 500 stock options data from 2004 to 2015. This indicates that option market developments have contributed to tighter spreads over time. In addition, when examining stock and ETF options separately, ETF options, a relatively

[6] More details refer to https://cdn.cboe.com/resources/government_relations/Understanding-Retail-Investors-Dynamic-Trading-Behavior-in-the-US-Options-Market_2024.pdf.

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

newer market segment, exhibit an even tighter percentage quoted spread of 0.1151, compared to the daily average percentage quoted spread of 0.1648 for stock options.

2.2 Price Improvement

Since wholesalers commit to providing price improvements (PI) to SLAN traders, we quantify the magnitude of these improvements for each individual trade using the following definition:

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

where P_{Actual} denotes the actual execution price of the trade, P_{Ask} and P_{Bid} represent the quoted ask and bid prices at the time of the trade, respectively, and the quoted spread is defined as the difference between the ask and bid prices ($P_{Ask} - P_{Bid}$). Intuitively, PI captures the fraction of the quoted spread that traders save due to wholesalers' price improvements. A higher value indicates that traders execute buy orders at prices closer to the bid and sell orders at prices closer to the ask, reflecting less spread revenue for wholesalers. When the actual execution price equals the midpoint, price improvement is 50%, indicating that investors trade at the "fair price".

Table 2 presents PI statistics calculated using the true trade direction from our proprietary PHLX dataset. Panel A shows that PI is relatively stable across different trade categories, with buy and sell trades receiving average PIs of 43.5% and 43.0%, respectively. However, buy trades exhibit significantly higher PI than sell trades (t-stat = 66.59), indicating that wholesalers may favor option buyers to earn premiums from these transactions. Additionally, call options have an average PI of 43.1%, significantly lower than the 43.5% observed for put options (t-stat = -53.70).

To better understand PI, we define the dollar PI as the numerator of equation (1). Panel B shows that the distribution of dollar PI is right-skewed, indicating that many trades receive relatively modest improvements in dollar terms. More than half of the transactions obtain dollar PI of two

cents or less. This result reflects constraints imposed by the one-cent minimum tick size in options markets, which restrict wholesalers from offering greater PI; unlike stock markets, the options market does not allow sub-penny pricing. These findings highlight that although percentage PI may appear sizable, they often result from very tight quoted spreads, creating a potentially misleading impression of low liquidity when measured using percentage quoted spreads. For example, for out-of-the-money (OTM) options quoted in the range of [0, 0.05], the percentage quoted spread can be as large as 200%.

3 Failure of the Midpoint Rule

In this section, we provide robust evidence that the low accuracy rate of the midpoint rule persists across various intraday time intervals and applies consistently to both stock and ETF options. Furthermore, we demonstrate that this low accuracy is not driven by quote mismatching. Rather, even when quote mismatching is absent, misclassification under the midpoint rule primarily arises from wholesalers offering substantial price improvements, frequently exceeding half of the quoted spread.

3.1 Low Accuracy of the Midpoint Rule

Researchers commonly apply stock trade classification methods, including the midpoint rule and the approach proposed by Muravyev (2016), to classify option trades as buys or sells. Savickas and Wilson (2003) evaluate four such classification rules using a proprietary 1995 Chicago Board Options Exchange (CBOE) dataset: the midpoint rule, the Lee and Ready (1991) method, the Ellis et al. (2000) method, and the tick rule. They demonstrate that the midpoint rule, classifying trades executed above (below) the midpoint as "buy" ("sell") trades, achieves the highest accuracy at approximately 83%. Building upon this, Muravyev (2016) propose a refinement where trades executed exactly at the midpoint are classified based on the quote rule of the exchange where

the trade took place. Bryzgalova et al. (2023) further confirm that ticker-level order imbalances computed from these methods exhibit a correlation exceeding 94%.

However, the structure of the options market has undergone significant changes since the early 2000s. Price improvement auctions, including CBOE's Automated Improvement Mechanism, International Securities Exchange's Price Improvement Mechanism, and Boston Options Exchange's Price Improvement Period, were introduced during this period to enhance trade execution quality. Additionally, the Options Order Protection and Locked/Crossed Market Plan, analogous to Regulation NMS Rule 611 (the Order Protection Rule) in the stock market, was officially implemented on August 31, 2009. This regulation ensures that option trades are executed at prices no worse than the National Best Bid and Offer (NBBO). Due to these structural developments, the midpoint rule's historical accuracy in classifying option trades may have substantially deteriorated.

In this section, we empirically verify this hypothesis using our proprietary PHLX dataset, showing that the midpoint rule's accuracy has significantly declined, achieving no better than 60% in 2024. We further demonstrate that this lower accuracy persists across various intraday intervals and is consistently observed in both stock and ETF options.

We primarily focus on assessing the accuracy of the midpoint rule for two main reasons. First, all existing option trade classification methods can be viewed as variants of the midpoint rule. As Bryzgalova et al. (2023) demonstrate, order imbalances calculated using these different midpoint-based methods exhibit very high correlations. Therefore, if the midpoint rule's accuracy is low, it implies that other closely related methods will likely share a similar degree of inaccuracy. Second, OPRA currently distributes only the NBBO and trade messages to data vendors.[8] Consequently, data vendors relying solely on OPRA feeds do not have access to the exchange-level quote data necessary for employing more precise methods, such as the one proposed by Muravyev (2016).

The dashed curve in Figure 1 represents the daily accuracy rate of the midpoint rule from January 1, 2024, to December 31, 2024. As illustrated, the accuracy rate fluctuates around

[8]See Section 5.2 of the OPRA Plan for more details: https://cdn.opraplan.com/documents/OPRA_Plan.pdf.

approximately 55%, with its peak accuracy slightly above 60%. Figure 2 further depicts the intraday patterns of midpoint rule accuracy for the entire sample as well as separately for stock and ETF options. The overall pattern reveals a decreasing accuracy trend throughout the trading session. Moreover, stock options consistently exhibit higher classification accuracy than ETF options. Given that stock options constitute about 80% of total SLAN trades, the aggregate accuracy largely mirrors the trajectory observed in stock options.

Additionally, option trading activity is unevenly distributed across the trading day, with notably higher volumes occurring during the first hour after market open. The midpoint rule accuracy correlates with this pattern, showing higher accuracy during periods of increased trading intensity. This relationship can be explained by intraday variations in trading dynamics. During peak trading intervals (particularly the first two half-hour intervals), heightened activity facilitates natural counterparty matching, allowing wholesalers to execute trades with minimal PI. In contrast, during off-peak periods, reduced trading intensity prompts wholesalers to offer greater PI to attract order flow. Consequently, when the offered PI surpasses 50% of the quoted spread, trades are misclassified by the midpoint rule.

To determine whether intraday accuracy of the midpoint rule is driven by trading intensity, we perform the following regression:

$$\text{Accuracy}_{i,t} = \beta_0 + \beta_1 \text{Spread}_{i,t} + \beta_2 \log(\text{Volume})_{i,t} + \beta_3 \text{VXX}_{i,t} + \text{Fixed Effects}_{i,t} + \varepsilon_{i,t}, \qquad (2)$$

where $\text{Accuracy}_{i,t}$ denotes the midpoint rule accuracy rate for the ith half-hour interval on day t; $\text{Spread}_{i,t}$ is the volume-weighted average dollar-quoted spread within the same half-hour interval; $\log(\text{Volume})_{i,t}$ is the logarithm of the trading volume; and $\text{VXX}_{i,t}$ is the midpoint quote price of the VXX ETF (proxying for VIX) at the end of the ith half-hour interval on day t.

Table 3 shows that, overall, higher option volume and elevated volatility (VXX) are associated with improved midpoint rule accuracy. However, distinct patterns emerge when the sample is separated into stock options and ETF options. For stock options, increased volatility negatively

impacts accuracy, and the quoted spread shows no statistically significant effect. In contrast, for ETF options, both higher volatility and wider quoted spreads are positively associated with improved accuracy. Despite these differences, both subsamples consistently indicate that higher trading volumes enhance midpoint rule accuracy. These findings align with the intraday patterns depicted in Figure 2, which exhibit higher accuracy during periods of peak trading intensity.

3.2 Minimal Impact of Quote Mismatching

The midpoint rule classification relies heavily on the midpoint price, which is computed from the bid and ask prices. Consequently, inaccuracies or mismatches in quoted prices can substantially undermine the performance of the midpoint rule. For example, Holden et al. (2023) show that adjusting for quote mismatches increases the accuracy of the Lee and Ready (1991) trade classification algorithm in the stock market from 86% to 92%. In this section, we provide three distinct pieces of evidence indicating that significant quote mismatches do not exist in our dataset. Therefore, the observed low accuracy of the midpoint rule is primarily attributable to the large PI offered by wholesalers, rather than inaccuracies stemming from quote mismatches.

To confirm the minimal existence of quote mismatches, we first verify the consistency between auction and trade timestamps in our dataset. After a PHLX member submits a SLAN order on behalf of their customers, an auction message is recorded in the PHLX TotalView dataset and broadcasted 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 confidential responses not visible to others. Once the auction concludes, the recorded trade is sent to OPRA.[9] We match SLAN trades recorded on OPRA with corresponding SLAN auction messages from the PHLX TotalView dataset. Consequently, the expected timestamp difference between OPRA and PHLX records should be approximately 100 milliseconds.

Figure 3 presents the time series of timestamp differences between OPRA and PHLX. On

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

average, the daily median timestamp difference is 99.54 milliseconds, indicating an average latency of 0.46 milliseconds ($= 100 - 99.54$) between OPRA and PHLX records. Additionally, the average 5th percentile of timestamp differences is 99.08 milliseconds, implying that fewer than 5% of SLAN trades experience latencies exceeding 0.92 milliseconds ($= 100 - 99.08$). A median timestamp difference slightly below 100 milliseconds suggests a minor delay in recording SLAN auction messages in the PHLX TotalView dataset. This interpretation is supported by the notable spike on November 11, 2024, when Nasdaq upgraded its NY11 data center in Carteret, NJ, and opened access.[10] Following this upgrade, the timestamp difference increased from 99.53 to 99.58 milliseconds, reflecting a reduction in latency from 0.47 to 0.42 milliseconds. Given that OPRA timestamps are recorded to the nearest millisecond, these negligible latencies suggest minimal quote mismatching issues. Given that midpoint rule accuracy critically depends on aligning trade prices with contemporaneous quotes, the minimal latency (less than one millisecond) between OPRA and PHLX data strongly suggests negligible quote mismatching issues. Thus, we expect this latency to have a minimal impact on midpoint classification accuracy.

A second piece of evidence indicating that quote mismatching has minimal impact on midpoint rule accuracy involves examining whether accuracy significantly improves when quotes from alternative timestamps around trades are used. If low accuracy primarily arises from quote mismatches, employing quote data from alternative time points should substantially enhance accuracy. Panel A of Table 4 reports the average accuracy of the midpoint rule for stock and ETF options using quotes from several time points: one second before the trade, 100 milliseconds before the trade, the last quote before the trade, the contemporaneous OPRA quote, the next quote after the trade, and one second after the trade. The results show that midpoint rule accuracy remains consistently low, ranging narrowly around 55% across all timestamps considered. The highest accuracy is observed for the contemporaneous quote (56.51%), closely followed by the quote immediately after the trade (56.47%).These figures are significantly lower than the accuracy

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

of over 80% typically observed in the stock market. Thus, the consistently low accuracy regardless of the timing of quote selection strongly suggests that the midpoint rule's poor performance in the options market is not primarily driven by mismatched quotes, but rather stems from other factors such as large PI offered by wholesalers.

A final piece of evidence against the quote mismatch hypothesis is based on the intuition that trades with slower quote updates should exhibit higher midpoint rule accuracy if quote mismatches were indeed the primary cause of low accuracy. To test this, we divide our sample into two equal groups according to the time difference between the last quote update and the recorded trade on OPRA, where a larger time difference indicates slower quote updates. Contrary to expectations, the group with slower quote updates (larger time differences) exhibits significantly lower midpoint rule accuracy, further rejecting the quote mismatch explanation.[11] In summary, advancements in exchange technology have substantially reduced latency between the exchange and OPRA, minimizing the potential for quote mismatches. Our analyses, including matching trades with quotes at multiple time points and subsample tests based on quote update speed, consistently reject the hypothesis that quote mismatching explains the midpoint rule's low accuracy. These findings collectively indicate that the decreased accuracy of the midpoint rule primarily results from wholesalers increasingly offering large PI on SLAN trades.

4 Machine Learning Identification Model

In this section, we introduce a machine learning framework designed to accurately recover the true trading direction of SLAN trades. We propose the GS-LASSO model, which integrates SHAP-LASSO feature selection with an Extreme Gradient Boosting (XGBoost) classification algorithm. After introducing the model, we evaluate its performance throughout the entire year of 2024 using a one-week training sample and demonstrate its substantial improvement over the commonly used

[11] Table 5 shows that the slower quote updates group (above median) achieves a 53.80% accuracy for the midpoint rule, whereas the faster quote updates group (below median) achieves 56.79%.

midpoint rule. Finally, we assess the robustness of the GS-LASSO model under various model configurations—varying the feature set, training window size, and retraining strategy—and show that its predictive accuracy remains consistently high across different specifications.

4.1 GS-LASSO Model

In this subsection, we introduce our GS-LASSO model. We begin by explaining the SHAP (SHapley Additive exPlanations) methodology and how it can be combined with LASSO regularization to objectively identify the most important features from a high-dimensional dataset. We then present the XGBoost algorithm, highlighting its strength in handling high-frequency financial data and capturing complex, non-linear interactions.

4.1.1 Feature Selection via SHAP-LASSO

SHAP, proposed by Lundberg and Lee (2017), is a widely adopted method for interpreting complex machine learning models by assigning importance values to each feature for every observation. It provides a unified framework that satisfies key interpretability properties and decomposes predictions from non-linear models into a linear additive sum of individual feature contributions. While many machine learning algorithms lack robust feature selection mechanisms, we address this limitation by introducing the SHAP-LASSO approach, which applies LASSO regularization to SHAP values to objectively identify the most influential features in our dataset.

SHAP provides a powerful framework for interpreting complex machine learning models by attributing the output of a model to its input features. Let $f(\mathbf{x})$ be a predictive model with N input features $\{x_1, x_2, \ldots, x_N\}$. For a given observation k and feature x_i, the Shapley value $\phi_i^{(k)}(f)$ quantifies the marginal contribution of x_i to the model's prediction $\hat{f}(\mathbf{x})$. It is computed by

averaging the change in prediction when x_i is added to all possible subsets of the other features:

$$\phi_i^{(k)}(f) \;=\; \sum_{S\subseteq\{1,\ldots,N\}\setminus\{i\}} \frac{|S|!\,(N-|S|-1)!}{N!}\Big[\hat{f}\big(x^{(k)}_{S\cup\{i\}}\big) - \hat{f}\big(x^{(k)}_{S}\big)\Big]. \tag{3}$$

where S denotes any subset of features excluding x_i, $|S|$ is the number of features in S, and $\hat{f}\big(x^{(k)}_S\big)$ represents the model's prediction based solely on the features in S for the k-th observation. This formulation captures the "fair share" of each feature in explaining the prediction and transforms the complex, often nonlinear model behavior into a linear additive form.

In binary classification tasks like ours, the model's prediction for each observation k can be expressed in terms of log-odds as:

$$\text{logit}(p_{(k)}) \;=\; \beta_0 + \sum_j \phi_j^{(k)}, \tag{4}$$

where $p_{(k)}$ is the predicted probability, β_0 is the baseline log-odds, and $\phi_j^{(k)}$ denotes the SHAP value of feature j for observation k. This decomposition enables interpretation at the individual feature level, making it particularly useful for understanding how predictions are formed.

While SHAP values are commonly visualized using importance plots to rank features, these rankings are often selected manually and lack objective feature selection. To address this, we propose a SHAP-LASSO method that leverages SHAP's linear additive property by applying LASSO logistic regression to the SHAP values. Specifically, we first train a machine learning model using the full feature set and compute SHAP values for over 33,000 out-of-sample observations. We then run a logistic regression of the binary outcomes on the SHAP values, using an ℓ_1-penalty to shrink irrelevant feature coefficients toward zero. The penalty parameter is selected via cross-validation, removing the need for arbitrary thresholds or manual adjustments.

This SHAP-LASSO framework provides a principled and automated way to identify the most informative features. By filtering out noise and retaining only the most impactful predictors, it improves both model interpretability and out-of-sample predictive accuracy.

4.1.2 Prediction via XGBOOST

XGBoost is a powerful and efficient implementation of the gradient boosting framework, widely used for predictive modeling in large-scale, complex datasets. The core idea of gradient boosting is to build an ensemble of decision trees sequentially, where each new tree is trained to correct the errors of the combined predictions from the previous trees. Specifically, at each iteration, the model fits a new tree to the negative gradient (i.e., the direction of steepest descent) of the loss function with respect to current predictions.

Formally, let $\hat{y}_i^{(t)}$ denote the predicted output for observation i after t iterations, and let $f_k(\mathbf{x}_i)$ represent the prediction of the k-th tree for input features $\mathbf{x}_i$. The model aggregates the predictions of all previous trees:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(\mathbf{x}_i),$$

where each f_k is a regression tree.

At each step, the algorithm selects the next tree f_t to minimize the following regularized objective function:

$$\text{Obj} = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t),$$

where $L(\cdot)$ denotes the loss function (e.g., logistic loss for classification), and $\Omega(f_t)$ is a regularization term that penalizes model complexity:

$$\Omega(f_t) = \gamma T + \tfrac{1}{2}\lambda \sum_{j=1}^{T} w_j^2.$$

Here, T is the number of leaf nodes in the tree, w_j is the output weight of the j-th leaf, and γ and λ are regularization parameters that help prevent overfitting.

XGBoost incorporates several enhancements—such as parallelized tree construction, sparse-aware learning, and efficient memory usage—that make it particularly well-suited for high-dimensional financial data, where complex, nonlinear interactions are common. For instance,

Bogousslavsky et al. (2024) use XGBoost to model interactions between informed trading, volume, and volatility, while Bartlett and O'Hara (2024) apply it to detect hidden liquidity and improve trade execution by predicting which venues offer the best price improvement.

Building on the strengths of XGBoost and SHAP, we propose a structured GS-LASSO framework that integrates SHAP-based interpretability with LASSO feature selection to enhance model efficiency and accuracy. The procedure unfolds in three key steps, involving two separate rounds of model training:

- **Step 1: SHAP Value Extraction from Initial Model.** We begin by training an XGBoost model using the full set of features on a baseline training sample, denoted as *Training Sample #1*. From this model, we compute SHAP (SHapley Additive exPlanations) values, which decompose the predicted probabilities into additive feature-level contributions for each observation.

- **Step 2: Feature Selection via LASSO on SHAP Values.** Using the SHAP values computed in Step 1, we apply LASSO logistic regression to select the most important features. The ℓ_1-penalty shrinks the coefficients of weak predictors toward zero, yielding a sparse, interpretable set of features that drive the model's predictions.

- **Step 3: Final Model Training with Selected Features.** With the reduced feature set from Step 2, we retrain an XGBoost model on *Training Sample #2*. This produces our final **GS-LASSO model**, which combines the non-linear predictive power of XGBoost with the interpretability and parsimony enabled by SHAP and LASSO.

By combining XGBoost's robust non-linear modeling capabilities with SHAP's interpretability and LASSO's feature selection power, the GS-LASSO framework efficiently filters out noisy features while preserving predictive strength and interpretability.

4.2 Empirical Performance

In this subsection, we evaluate the performance of our GS-LASSO model using data from the first week of 2024 as both *Training Sample #1* and *Training Sample #2*. Remarkably, we find that training the model on just five trading days enables it to consistently predict trading directions with significantly higher accuracy than the midpoint rule throughout the remainder of the year. This result highlights the model's strong generalizability and robustness, even when trained on a limited initial sample.

4.2.1 Feature Selection

We initially construct 50 features (listed in Table A1) that capture a wide range of information relevant to option trades. These features fall into three categories. The first group consists of 21 option trade-specific features, including trade size, trade price, option moneyness, midpoint rule direction, effective spread, and others. The second group captures time-series dynamics around option trade execution, such as return and volatility computed over intervals of 100 milliseconds, 1 second, and 5 seconds. The final group incorporates information from the stock market, including the contemporaneous price of the VXX ETF (a proxy for the VIX index), and similar time-series metrics for the underlying stock.

In our baseline model, we use the first week of data—that is, the first five trading days of 2024—for initial model training and feature selection (*Training Sample #1*). The robustness of this selection is further discussed in Section 4.3. Applying the first three steps of our SHAP-LASSO framework, we reduce the original 50 features to 36 key predictors. Among the selected features, dollar quoted spread, trade size, midpoint rule direction, and midpoint rule effective spread stand out as particularly influential,

Table A1 presents the features selected by our GS-LASSO model, along with their SHAP values ranked by both maximum absolute and mean contributions. Based on maximum absolute

SHAP values, the top five features are: Dollar Quoted Spread, Trade Size, Option Return over $[t-1\text{s}, t]$, Midpoint Rule Direction, and Option Volatility over $[t-100\text{ms}, t]$. When ranked by mean SHAP values, the top contributors are: Moneyness, Percentage Quoted Spread, Effective Spread, Dollar Quoted Spread, and Time Quote Difference. The consistent ranking of Dollar Quoted Spread across both metrics highlights its strong and stable importance in classification.

However, our GS-LASSO model goes beyond simple SHAP-based rankings by accounting for interactions and joint explanatory power that may not be evident from univariate SHAP scores alone. While SHAP offers a powerful diagnostic tool for interpreting model output, using it in isolation can lead to the omission of economically relevant but subtle predictors. For example, features such as the Security Type indicator (distinguishing between stock and ETF options) or market-wide volatility (VXX) may not appear among the top SHAP-ranked features, yet are selected by LASSO due to their added predictive value when combined with other features. This richer modeling framework allows GS-LASSO to capture economic mechanisms—such as the observed decline in midpoint rule accuracy during volatile periods—by incorporating relevant but less obvious features like VXX. In this way, GS-LASSO delivers a more comprehensive and interpretable feature set than would be achieved through SHAP analysis alone.

4.2.2 Prediction Accuracy

We continue to use the first week of data as the training sample for model prediction in Step 3 (*Training Sample #2*). Figure 4 plots the time series of classification accuracy for our GS-LASSO XGBoost model versus the midpoint rule over the full year of 2024. The figure shows that our model consistently delivers substantially higher accuracy than the midpoint rule throughout the sample period. On average, the GS-LASSO model outperforms the midpoint rule by approximately 20 percentage points. While the midpoint rule's accuracy fluctuates between 45% and 60%, the GS-LASSO model typically maintains an accuracy between 70% and 80%. This persistent and sizable gap highlights the robustness and effectiveness of our machine learning approach relative

to traditional classification methods.

To better understand the performance of our GS-LASSO model and the midpoint rule under different market conditions, we divide the full sample into subsamples based on key trade characteristics. Specifically, we partition the data into equal-sized groups using the median values of four key features: trade price, trade size, dollar quoted spread, and midpoint effective spread. For each subsample, we calculate the average classification accuracy of both models. As reported in Table 5, the GS-LASSO model consistently outperforms the midpoint rule across all subsamples, with an average improvement of approximately 20 percentage points. This highlights the robustness of our model in capturing trade direction across diverse trading environments.

In particular, for trades involving low-priced options (below-median price group), the midpoint rule performs poorly, achieving only 51.71% accuracy, compared to 58.91% in the above-median group. This suggests that wholesalers are more likely to offer substantial PI in low-priced trades, complicating midpoint-based classification. By contrast, our GS-LASSO model increases the accuracy in this low-price group by 28.31%, effectively correcting these misclassifications. A similar pattern is observed for trades with low dollar quoted spreads, where our model improves classification accuracy by 27.33%, from 54.43% to 81.76%. While the GS-LASSO model delivers improvements across the board, the magnitude of improvement differs across subsamples, indicating that incorporating feature-specific sub-models tailored to market microstructure characteristics could further enhance classification performance.

4.3 Model Robustness

We now assess the robustness of our GS-LASSO model along two key dimensions. First, we examine the sensitivity of the selected features to the choice of the initial training sample (*Training Sample #1*) used for feature selection. Second, we evaluate the model's predictive stability across different validation periods (*Training Sample #2*) used for re-estimation and out-of-sample testing.

To evaluate the robustness of feature selection, we extend our analysis beyond the baseline

model, which uses only the first week of data for *Training Sample #1*. Specifically, we repeat the SHAP-LASSO procedure for all 50 weeks of data in our sample year. For each week, we retrain the initial XGBoost model, compute SHAP values, and apply LASSO to identify the most predictive features. This rolling analysis allows us to observe whether certain features are persistently selected across different training periods or if the selected set is highly sensitive to the time window chosen.

As shown in Figure 5, the selected features exhibit a high degree of stability. Most option trade-specific features, such as trade size, dollar quoted spread, and midpoint rule direction, are consistently chosen across weeks. Additionally, time-series features constructed from option prices—such as short-term return and volatility around the trade timestamp—are selected more frequently than comparable features based on the underlying stock. This finding confirms that option-level dynamics contain stronger signals for classifying trade direction than stock-level metrics.

To assess the robustness of our GS-LASSO model's predictive performance, we evaluate a comprehensive set of model configurations that vary by feature set, training sample size, and retraining strategy. Specifically, we compare three groups of XGBoost models: (1) a baseline model using 21 option-specific features from SLAN; (2) a full-feature model incorporating all 50 features, including time-series information and market-level metrics; and (3) our GS-LASSO model using a refined subset of 36 features selected via the SHAP-LASSO procedure.

Within each group, we train models using three different samples (*Training Sample #2*)—one day, one week (five trading days), and one month (22 trading days)—to assess performance across varying data depths. We also consider two retraining strategies: a static approach that uses a fixed training window and a rolling approach that continuously updates the training set by incorporating the most recent data while discarding the oldest. This yields a total of 18 XGBoost model specifications. Our baseline model introduced in Section 4.2.2 corresponds to the one-week static configuration of group (3), the GS-LASSO model.

As reported in Table 6, all 18 model configurations consistently outperform the midpoint rule by a wide margin. Regardless of sample size or retraining strategy, each model achieves significantly higher out-of-sample classification accuracy, confirming the robustness of our machine learning framework. This performance advantage holds across both static and rolling schemes and demonstrates the resilience of our model to variations in implementation.

Among these configurations, the GS-LASSO models—those trained on SHAP-LASSO selected features—consistently deliver superior performance. Notably, the one-week rolling GS-LASSO model achieves the highest average out-of-sample accuracy of 75.94% over the testing period from January 9 to December 31, 2024. This highlights the advantage of combining targeted feature selection with adaptive model updating to capture evolving patterns in trade behavior.

Table 7 compares the performance of the six GS-LASSO model specifications, varying by training sample size (one day, one week, or one month) and by retraining strategy (static vs. rolling). The one-week rolling model achieves the highest average accuracy, and as shown in the first row, it significantly outperforms each of the other five configurations. Among models that use SHAP-LASSO selected features, those employing a rolling retraining strategy consistently outperform their static counterparts, aligning with our expectations that rolling models better capture evolving market dynamics. Notably, even with just one week of static training data, our baseline model (Section 4.2.2) already delivers strong performance. Exploring alternative specifications confirms the robustness of our approach and highlights the model's potential for further improvements in predictive accuracy under more adaptive or data-rich settings.

5 When Do Wholesalers Provide Large Price Improvements?

Our GS-LASSO model significantly enhances the accuracy of trade direction classification and reveals that wholesalers offer large and more frequent PI than previously recognized. Leveraging the true trade direction data from PHLX, this section explores two key questions: (1) Do

wholesalers generate profits after providing high PI? (2) What types of SLAN trades are more likely to receive higher PI from wholesalers?

Figure 6 presents the empirical cumulative distribution function of PI based on our dataset of 14.45 million SLAN trades. The data show that only 17.84% of trades receive no PI, while 11.80% receive full PI—i.e., investors buy at the bid and sell at the ask. Another 7.56% of trades occur at the midpoint, receiving a 50% PI, resulting in a distinct jump in the distribution around that point. These patterns highlight the strategic discretion wholesalers exercise in offering PI. Moreover, there are notable differences between stock and ETF options: 15.38% of stock option trades receive no PI compared to 28.53% for ETF options, while 9.41% of stock and 22.18% of ETF trades receive full PI. Notably, 30.77% of total trading volume benefits from at least a 50% PI—levels that imply wholesalers incur losses on the quoted spread. These observations underscore the importance of understanding both the profitability and selection mechanism behind wholesalers' PI strategies.

To evaluate whether wholesalers profit after providing PI, we assume that they either close or hedge their option positions at the subsequent quote midpoint after a fixed time interval. For each SLAN trade i at time t, we compute the dollar return to the wholesaler as follows:

$$\begin{aligned} Ret_{i,t} &= -100 \cdot \text{Dir}_{i,t} \cdot \text{Volume}_{i,t} \cdot \left(P_{\text{midpoint},i,t+\tau} - P_{\text{actual},i,t}\right) \\ &= 100 \cdot \text{Dir}_{i,t} \cdot \text{Volume}_{i,t} \cdot \left(P_{\text{actual},i,t} - P_{\text{midpoint},i,t+\tau}\right), \end{aligned} \tag{5}$$

where 100 is the standard multiplier for an option contract, $\text{Dir}_{i,t}$ is the true trade direction from the PHLX dataset (positive for buy trades and negative for sell trades), and $\text{Volume}_{i,t}$ is the number of contracts traded. $P_{\text{actual},i,t}$ is the transaction price after the investor receives PI, and $P_{\text{midpoint},i,t+\tau}$ is the midpoint of the quoted bid and ask prices τ seconds after the trade. We examine a range of τ values in our analysis: the next available quote, as well as fixed intervals of 1, 5, 15, 30, 60, 90, and 120 seconds.

Table 8 presents wholesalers' profits under various assumptions about when they close or hedge their inventory, specifically at the quote midpoint after different time intervals (τ) following each SLAN trade. Regardless of the assumed time horizon, wholesalers consistently generate positive returns. If they close positions at the next available quote, they earn an average daily profit of \$1.29 million. Even with a more conservative assumption—waiting 120 seconds to close positions—they still earn a substantial average daily profit of \$0.58 million. This persistent profitability across all settings highlights the effectiveness of wholesalers' trade selection and inventory management strategies. While profits decline slightly with longer holding intervals, this trend is expected and consistent with findings in the stock market (Conrad and Wahal (2020)).

Across the full sample, the proportion of profitable trades remains robust, with over 50% of trades yielding gains even at the 120-second horizon. Notably, when profits are weighted by trade volume, the share of profitable volume exceeds 70%, underscoring that wholesalers are more likely to profit on larger trades. The persistent gap between count-based and volume-weighted profit rates suggests that wholesalers successfully identify and execute larger, more profitable trades—ensuring sustained gains even when only a modest majority of trades are winners.

We further decompose the wholesaler's dollar return in Equation (5) into two components: spread profit and position profit, as shown below:

$$\begin{aligned}\text{Ret}_{i,t} &= 100 \cdot \text{Dir}_{i,t} \cdot \text{Volume}_{i,t} \cdot \left(P_{\text{actual},i,t} - P_{\text{midpoint},i,t+\tau}\right) \\ &= \underbrace{100 \cdot \text{Dir}_{i,t} \cdot \text{Volume}_{i,t} \cdot \left(P_{\text{actual},i,t} - P_{\text{midpoint},i,t}\right)}_{\text{Spread Profit}} + \underbrace{100 \cdot \text{Dir}_{i,t} \cdot \text{Volume}_{i,t} \cdot \left(P_{\text{midpoint},i,t} - P_{\text{midpoint},i,t+\tau}\right)}_{\text{Position Profit}}.\end{aligned} \tag{6}$$

Intuitively, the spread profit reflects the immediate gain (or loss) from executing the trade at a favorable price relative to the prevailing midpoint at the time of the trade—essentially capturing how much PI the wholesaler concedes or retains. The position profit, on the other hand, measures the gain (or loss) from holding the acquired position over a short horizon until it is closed or hedged, based on changes in the option's quoted midpoint. This decomposition allows us to separately

assess whether wholesalers profit from favorable execution or from strategic position-taking based on expectations of short-term price movements.

Table 9 presents the decomposition of wholesalers' profits into spread and position components, as defined in Equation (6). First, we find that the majority of profit comes from the spread component, which is locked in at the time of execution and unaffected by subsequent price movements. On average, wholesalers earn a daily spread profit of $1.33 million.

Second, position profit—the return from holding the option after execution—is consistently negative across all time intervals, declining from -$0.36 million when positions are closed at the next quote to -$0.75 million after 120 seconds. While modest relative to spread profits, these losses suggest that wholesalers could potentially reduce risk or enhance profitability by hedging more quickly. The negative position profit also indicates that many SLAN trades may be informed, as wholesalers lose value on positions after execution.

Third—and most importantly—when we split trades by the size of PI, we observe a striking contrast: trades receiving at least 50% PI generate positive position profits, while those receiving less than 50% PI yield negative ones. This pattern strongly suggests that wholesalers are able to distinguish informed from uninformed traders. For potentially informed trades, wholesalers offer minimal PI and seek to mitigate losses through tighter execution. In contrast, for likely uninformed trades, they are willing to provide larger PI, confident that they can profit from the position. This behavior underscores the economic rationale behind wholesalers' selective provision of PI and highlights their sophisticated risk management in balancing spread revenue and position exposure.

To examine which types of SLAN trades receive higher PI from wholesalers, we estimate the following regression model:

$$\begin{aligned} \mathrm{PI}_{i,t} = {} & \beta_0 + \beta_1 \mathrm{cp}_{i,t} + \beta_2 \mathrm{size}_{i,t} + \beta_3 \mathrm{price}_{i,t} \\ & + \beta_4 \mathrm{buyind}_{i,t} + \beta_5 \mathrm{yexp}_{i,t} + \beta_6 \mathrm{bs_spread}_{i,t} + \beta_7 \mathrm{tqdiff}_{i,t} \\ & + \alpha_i + \gamma_t + \varepsilon_{i,t}, \end{aligned} \tag{7}$$

where $\text{PI}_{i,t}$ is the percentage price improvement for trade i on day t, measured in percentage points. The explanatory variables are defined as follows: $\text{cp}_{i,t}$ is a dummy variable equal to 1 if trade i on day t is a call option and 0 otherwise; $\text{size}_{i,t}$ is the trade size; $\text{price}_{i,t}$ is the trade execution price; $\text{buyind}_{i,t}$ is a dummy equal to 1 for buy trades and 0 for sell trades; $\text{yexp}_{i,t}$ is the time to expiration (in years); $\text{bs_spread}_{i,t}$ is the dollar-quoted bid-ask spread; and $\text{tqdiff}_{i,t}$ denotes the trade–quote time difference, as defined in Table A1. We include security fixed effects, α_i, and date fixed effects, γ_t, to control for unobserved heterogeneity across securities and trading days.

Table 10 reports the regression results. The coefficient on the buy trade indicator is 0.5238, suggesting that wholesalers tend to offer an additional 0.52% PI for buy trades. The coefficient for the call option indicator is -0.3601, showing that call options receive 0.36% lower PIs compared to puts. The dollar quoted spread has a statistically significant coefficient of 0.8654, meaning that a one standard deviation increase in spread (0.59) is associated with only a modest 0.51% ($= 0.8654 \times 0.59$) increase in PI. These results collectively suggest that wholesalers provide more generous PI for buy orders, especially put options, and for trades with wider quoted spreads.

6 Conclusion

Due to the widespread use of price improvement auctions in today's options market, traditional stock-based trade classification rules—such as the midpoint rule—now perform poorly in correctly identifying trade direction. To address this challenge, we introduce the GS-LASSO model, which integrates the XGBoost machine learning algorithm with LASSO-based feature selection using SHAP values. Our model substantially improves classification accuracy, raising it from approximately 56% to nearly 76%. This methodological advancement enhances the reliability of empirical studies on trading costs, spreads, and execution quality. Moreover, it reveals that substantial price improvements can coexist with wholesaler profitability, as wholesalers strategically offer larger improvements to uninformed traders while effectively managing risk.

Our findings also carry broader regulatory implications. The U.S. options market, which has long operated under an order-by-order auction structure, offers a compelling parallel to the SEC's recent proposals aimed at increasing competition in the stock market. Our results suggest that retail investors stand to benefit from meaningful price improvements when market makers compete for individual orders, consistent with auction theory predictions of improved allocative efficiency.

In sum, this paper not only introduces a robust machine learning method for accurately classifying trade direction in retail option markets, but also delivers meaningful implications for policymakers and practitioners. As a promising avenue for future research, our framework can be extended to identify informed trading behavior, assess market quality across different trading venues, or evaluate the impact of regulatory changes on retail execution outcomes.

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.

Figure 1. Time Series of Midpoint Rule Accuracy Rate

The figure presents the time series of accuracy rate for the midpoint rule . Data period covers from January 1, 2024 to December 31, 2024.

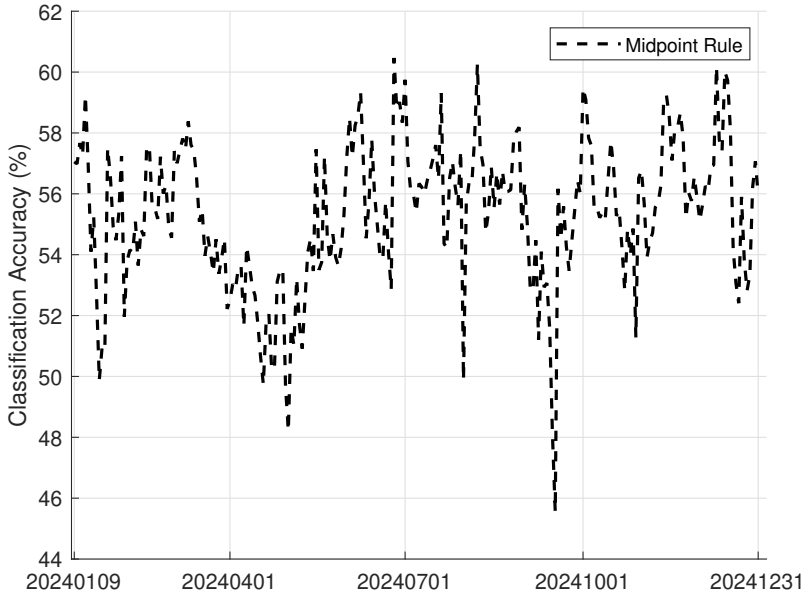


Figure 2. Intraday Accuracy Rate of Midpoint Rule

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

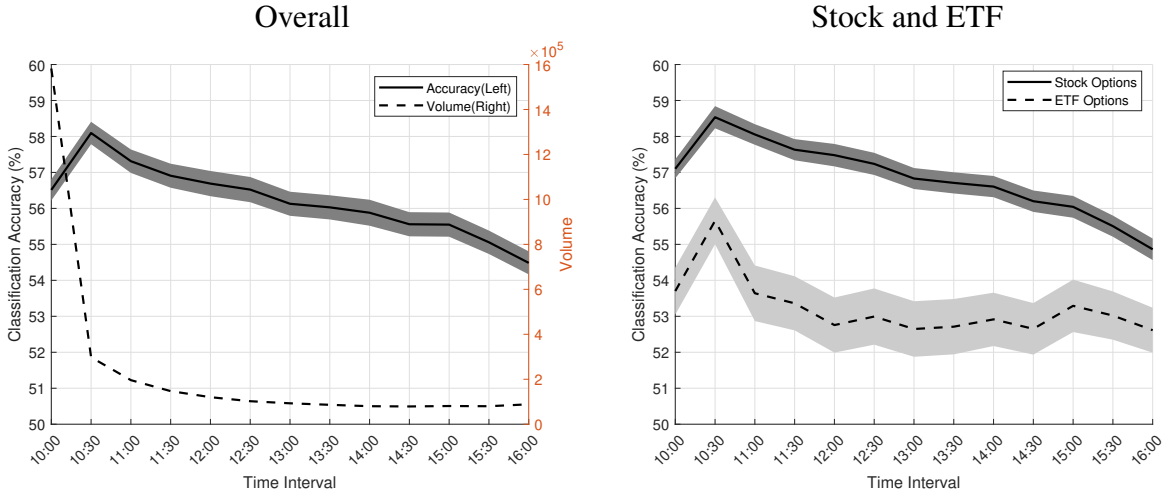


Figure 3. Duration of PHLX Auction in Data

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.

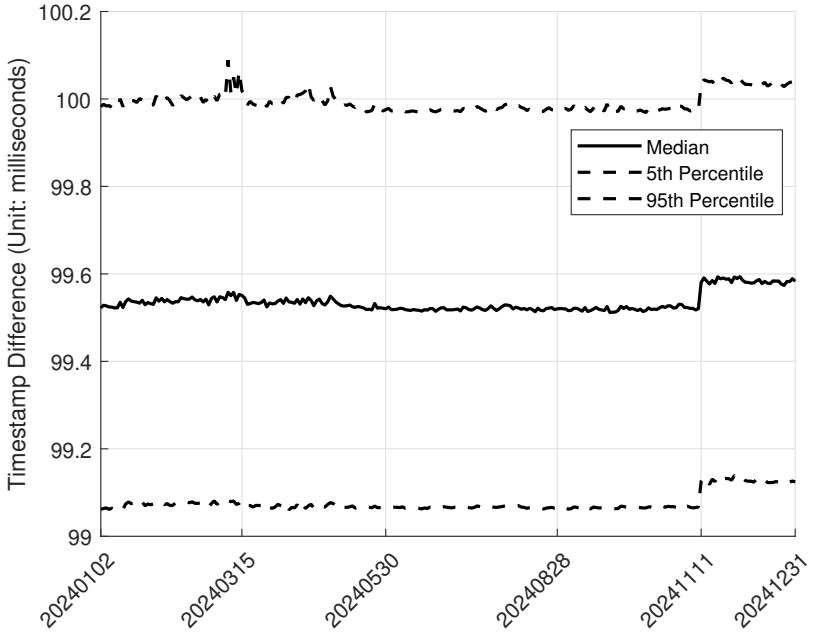


Figure 4. Comparison between GS-LASSO Model and Midpoint Rule

The figure presents the time series of accuracy rate for GS-LASSO Model and the midpoint rule, . Data period covers from January 1, 2024 to December 31, 2024.

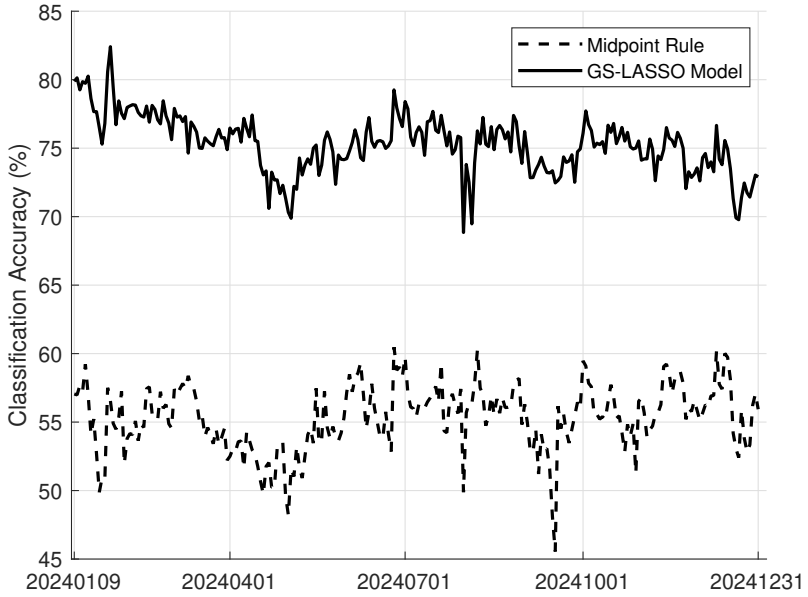


Figure 5. Feature Selection from SHAP-LASSO

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

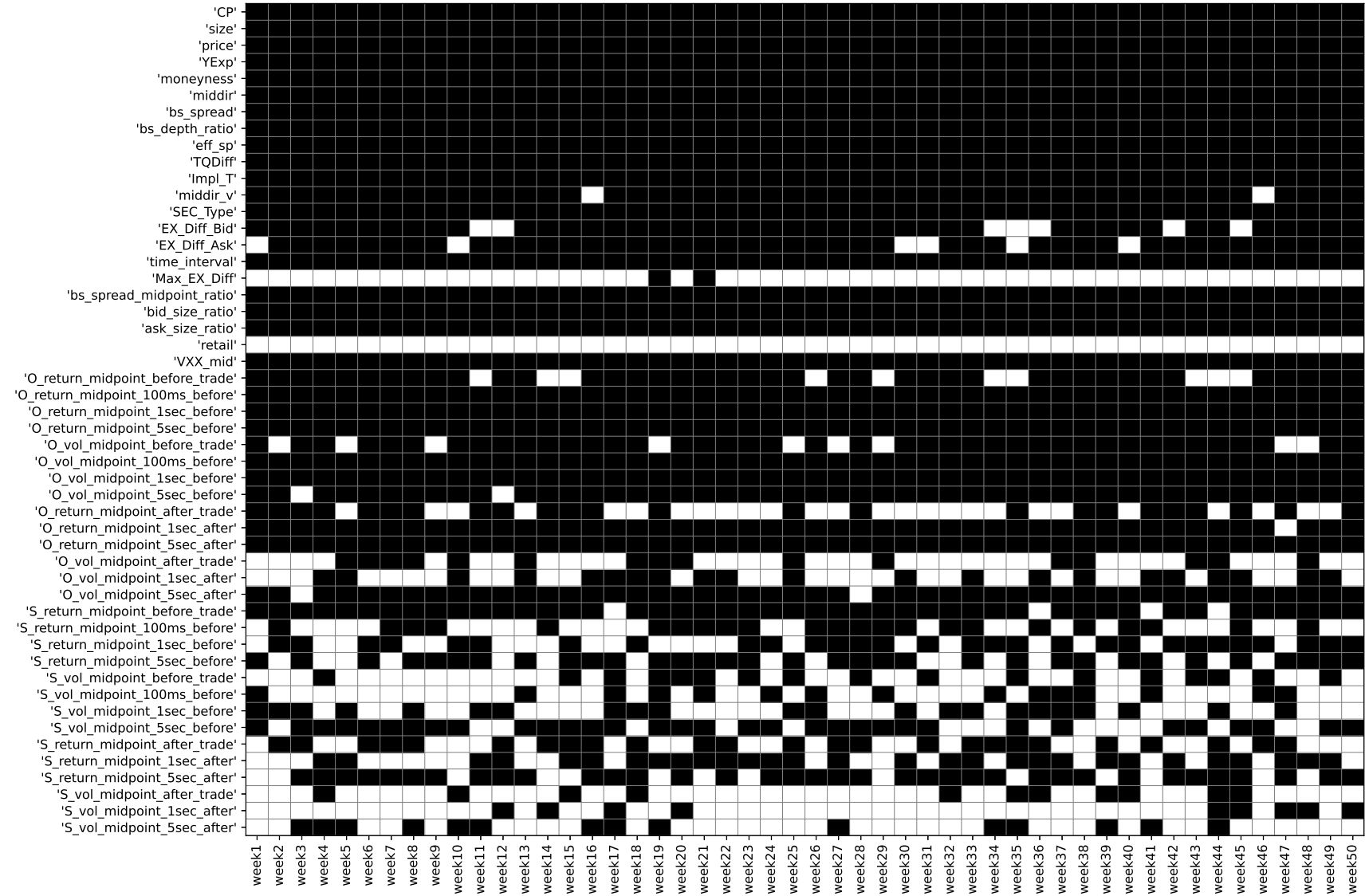
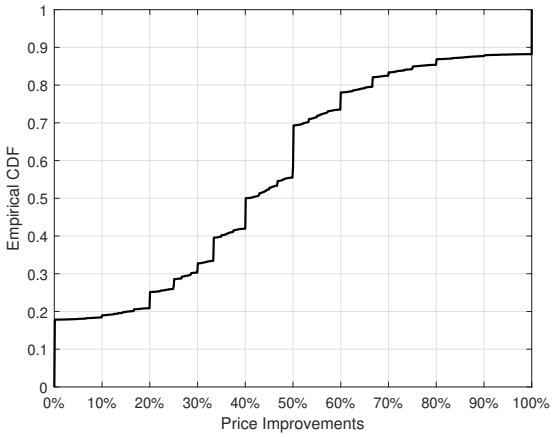


Figure 6. Empirical Cumulative Distribution Function of Price Improvements

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

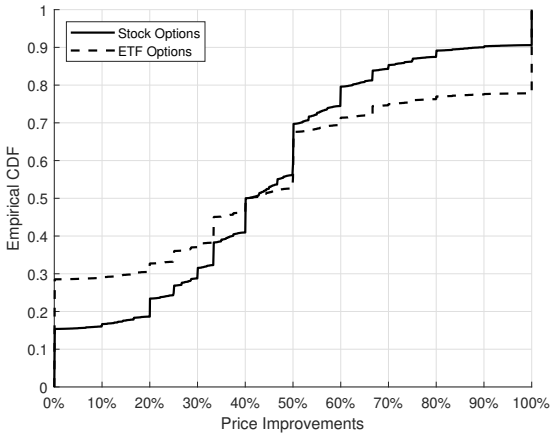


Table 1 Descriptive Statistics of Daily SLAN Trades

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 coefficient estimates of the intraday accuracy regression (2). 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 Model Accuracy Comparison Across Various Feature Subsamples

Table presents the various subsamples accuracy of two models: the GS-LASSO 1W Static model(our baseline model) and the midpoint rule. 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 baseline model 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	75.02%	56.44%	18.58%***	75.77%	53.67%	22.10%***
			(129.48)			(119.62)
Price	80.01%	51.71%	28.31%***	70.62%	58.91%	11.71%***
			(102.70)			(149.90)
YExp	78.19%	54.63%	23.56%***	72.46%	55.97%	16.49%***
			(114.20)			(131.19)
Moneyness	75.25%	55.14%	20.11%***	75.40%	55.46%	19.94%***
			(121.18)			(124.75)
Dollar Quoted Spread	81.76%	54.43%	27.33%***	67.92%	56.36%	11.55%***
			(98.06)			(122.01)
Bid-Ask Depth Ratio	75.19%	55.26%	19.92%***	75.46%	55.34%	20.12%***
			(123.11)			(127.88)
Midpoint ES	70.96%	47.18%	23.79%***	79.69%	63.44%	16.25%***
			(212.09)			(70.52)
Quote Trade Time Diff	77.85%	56.79%	21.06%***	72.79%	53.80%	18.99%***
			(122.41)			(124.99)

Table 6 Average Accuracy Across Different Modeling Approaches

This table presents the mean, standard deviation, minimum, and maximum accuracy (%) for three sets of models: SLAN Features Models, ALL Features Models, and GS-LASSO Models. Each model is tested in six approaches under three different time horizons (1 Day, 1 Week, 1 Month) and with Static vs. Rollingover approaches.Data period covers from January 1, 2024 to December 31, 2024.

Time point	Mean	Std	Min	Max
Panel A: SLAN Features Models				
1 Day Static	73.65%	2.17%	67.10%	80.38%
1 Day Rollingover	74.29%	2.01%	67.64%	80.84%
1 Week Static	73.80%	2.09%	67.41%	81.05%
1 Week Rollingover	74.48%	1.91%	68.11%	81.10%
1 Month Static	73.80%	1.86%	67.88%	77.94%
1 Month Rollingover	74.28%	1.80%	68.57%	78.23%
Panel B: ALL Features Models				
1 Day Static	75.21%	2.19%	68.66%	81.94%
1 Day Rollingover	75.70%	2.08%	69.31%	82.13%
1 Week Static	75.31%	2.10%	68.86%	82.43%
1 Week Rollingover	75.94%	1.91%	69.91%	82.34%
1 Month Static	75.34%	1.87%	69.36%	79.54%
1 Month Rollingover	75.74%	1.82%	70.05%	79.70%
Panel C: GS-LASSO Models				
1 Day Static	75.21%	2.19%	68.64%	81.92%
1 Day Rollingover	75.70%	2.08%	69.31%	82.06%
1 Week Static	75.29%	2.10%	68.84%	82.41%
1 Week Rollingover	75.94%	1.91%	69.87%	82.40%
1 Month Static	75.21%	2.19%	68.64%	81.92%
1 Month Rollingover	75.74%	1.82%	70.06%	79.61%

Table 7 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 8 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 9 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 10 Price Improvement Regression with Securities and Day Fixed Effects

This table displays coefficient estimations of a fixed effects regression model (7) to control for unobserved heterogeneity across both securities and time. 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

Internet Appendix

1 Features 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