Inferring Option Trade Direction with Machine Learning*

Si Gao
University of Texas at Dallas

Xiaoxiao Tang
University of Texas at Dallas

Guofu Zhou
Washington University in St. Louis

Zhaoque (Chosen) Zhou
University at Albany – SUNY†

First Draft: January 2025
Current Version: August 2026

*We thank Dmitriy Muravyev, Neil Pearson, Jing Wen, and Rezwan Hossain Ratul, as well as participants at the 2025 Financial Management Association Annual Meeting and the 2025 Annual Meeting of the Southern Finance Association and seminar participants at Chengdu University of Technology, Fudan University, the University at Albany, SUNY, and Washington University in St. Louis for helpful comments and suggestions.

†Si Gao, Jindal School of Management, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080-3021; phone: 314-445-9435 and email: si.gao@utdallas.edu. Xiaoxiao Tang, Jindal School of Management, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080-3021; phone: 972-883-4736 and email: xiaoxiao.tang@utdallas.edu. Guofu Zhou, John M. Olin Business School, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130-4899; phone: 314-935-6384 and email: zhou@wustl.edu. Zhaoque (Chosen) Zhou, Massry School of Business, University at Albany – SUNY, 1400 Washington Avenue, Albany, NY 12222; phone: 518-956-8304 and email: zzhou22@albany.edu.

Inferring Option Trade Direction with Machine Learning

Abstract

We provide the first direct evaluation of the widely used midpoint rule for identifying option trade direction in a modern option-auction setting using proprietary exchange data containing the true trade initiator. We show that the midpoint rule achieves an average daily accuracy of only 55.34%, because auction-based price improvement fundamentally weakens price-based trade signing. We develop GS-LASSO, an interpretable machine learning framework that treats trade direction as a state-dependent inference problem and increases out-of-sample classification accuracy to 80.46%. Our findings identify a major problem in modern option markets and provide a new framework for recovering trade direction from public data.

JEL Classification: G12, G14, C55, C58

Keywords: Trade Direction Classification; Options Market Microstructure; Price Improvement; Machine Learning; Auction-Based Trading

1 Introduction

Accurately identifying option trade direction is a fundamental task in market microstructure. It underlies the measurement of transaction costs, effective spreads, order imbalance, demand pressure, and price impact, and is essential for a wide range of empirical studies, including those on option order imbalance and return predictability (Hu, 2014; Zhan, Han, Cao, and Tong, 2022), retail option demand (Bryzgalova, Pavlova, and Sikorskaya, 2023), the effects of retail demand on the implied-volatility surface (Eaton, Green, Roseman, and Wu, 2026), and execution quality in option auctions (Hendershott, Khan, and Riordan, 2026). Although the widely used midpoint rule has long served as the standard approach for classifying option trades, its accuracy has never been directly validated using exchange-recorded trade direction in the modern U.S. options market.[1] Consequently, it remains unclear how reliable the midpoint rule is in option markets and whether its accuracy can be substantially improved.

In this paper, we address these questions and make four contributions. First, using proprietary Philadelphia Stock Exchange (PHLX) data containing the exchange-recorded initiator side, we provide the first direct evaluation of the midpoint rule in a modern option-auction setting.[2] We show that it achieves an average daily classification accuracy of only 55.34%, only slightly better than a random guess. The exchange-quote refinement of Muravyev (2016) improves the accuracy to 60.16%, but the performance remains unsatisfactory. Second, we develop Guided SHAP-LASSO (GS-LASSO), a machine-learning framework specifically designed for trade-sign classification. By combining model flexibility with interpretable variable selection, GS-LASSO increases the out-of-sample classification accuracy to 80.46%, representing a substantial

[1] Using proprietary CBOE data from July through December 1995, Savickas and Wilson (2003) report that the midpoint rule correctly classifies 83% of classifiable trades. Since then, however, the U.S. options market has undergone fundamental structural changes, including widespread multiple listing and intermarket linkage, decimalization and penny quoting, the transition to electronic and hybrid trading, and the adoption of price-improvement mechanisms. These developments have substantially altered quoting and execution practices, so the accuracy documented for the 1995 market may not generalize to the modern market.

[2] Two recent studies independently document similar limitations of the midpoint rule in options markets (Amaya, Garcia-Ares, Pearson, and Vasquez, 2025; Barardehi, Bogousslavsky, and Muravyev, 2025).

improvement over existing methods.

Third, despite its flexibility, the framework remains highly interpretable. It identifies publicly observable market-state variables that contribute to trade signing beyond the execution price, including liquidity, market depth, trade size, short-horizon price dynamics, and fair-value measures, thereby providing new economic insights into how price improvement varies with trading and market conditions. Fourth, we demonstrate that trade-direction errors have important implications for both empirical measurement and substantive inference. GS-LASSO produces order-imbalance measures that are substantially closer to true order imbalance and effective half-spread estimates that greatly reduce the transaction-cost overstatement induced by midpoint-based classification. We further show, in a same-sample gross-performance application, that inference about the gross timing component of SLAN trading can depend materially on the direction measure used.

We implement our GS-LASSO framework using a comprehensive dataset of retail-driven option trades from PHLX, matched with transaction and quote data from the Options Price Reporting Authority (OPRA), Trade and Quote (TAQ) data for the underlying stocks, and option-market variables from SpiderRock. Our final sample consists of 46.68 million single-leg auction, non-ISO (SLAN) option trades executed in 2024. A key advantage of these data is that the exchange records the true trade initiator, enabling us to evaluate trade-classification methods directly rather than relying on indirect proxies. This setting is also economically important. SLAN trades account for 63.10% of total option trading volume on PHLX during our sample period, indicating that auction-based retail execution represents a substantial and representative segment of today's listed-options market.

Our central empirical finding is that traditional price-based trade-signing rules perform poorly because price improvement is both pervasive and economically large. On average, price improvement equals 43.23% of the quoted bid–ask spread. Moreover, 13.86% of trades receive exactly 50% price improvement, placing the transaction price precisely at the quote midpoint and

eliminating the directional information contained in the execution price. An additional 30.67% of trades receive more than 50% price improvement, causing the transaction price to cross the midpoint and mechanically reversing the sign assigned by the midpoint rule. Consequently, 44.53% of SLAN trades fall into a region in which midpoint signing either loses its directional signal or assigns the wrong trade direction. Consistent with this mechanism, the midpoint rule achieves an average daily classification accuracy of only 55.34%, and its poor performance is remarkably persistent over time and across buy and sell orders, call and put options, and other market segments. These findings demonstrate that the midpoint rule fails not because of quote-matching errors or a small number of atypical contracts, but because auction-based price improvement fundamentally weakens the informational content of transaction prices.

Our proposed GS-LASSO addresses this measurement problem by recovering trade direction when the execution price alone no longer provides sufficient information. The key insight is that, in auction-based option markets, trade direction must be inferred from the broader market environment rather than from a single price-location rule. GS-LASSO first applies a low-complexity machine-learning model to extract stable nonlinear signals from publicly available information. SHAP values are then used to decompose the model's predictions into feature-level contributions, and LASSO identifies the predictors that retain incremental explanatory power after controlling for the others. Finally, a high-capacity XGBoost classifier is estimated using only the selected predictors. This three-stage design combines interpretability with predictive power. The initial model and LASSO screening filter out noisy and unstable predictors, while the final XGBoost model retains the flexibility required to capture nonlinearities, interactions, and infrequent but economically meaningful predictive patterns.

GS-LASSO substantially improves trade classification. In our baseline specification, it achieves an average out-of-sample classification accuracy of 80.46%, approximately 25 percentage points higher than the midpoint rule. The improvement is economically large and remarkably broad-based across option types and trading environments. More importantly, the gains are largest precisely in the settings where the midpoint rule performs worst. For example, for

options with below-median prices, GS-LASSO attains an accuracy of 84.25%, compared with only 51.87% for the midpoint rule. Similarly, for trades with below-median quoted dollar spreads, the corresponding accuracies are 85.00% and 54.74%, respectively. These findings indicate that GS-LASSO does not outperform the midpoint rule by exploiting opaque statistical regularities. Instead, it recovers economically relevant information that becomes particularly valuable when execution price loses its traditional directional content.

The economic sources of the improvement are also transparent. The selected predictors include measures of liquidity and market depth, trade size, short-horizon option and underlying-stock price dynamics, and SpiderRock fair-value and volatility-surface variables. Collectively, these variables determine trade direction in a way that cannot be summarized by a single price threshold. The execution price therefore remains informative, but only conditionally: its informational content depends critically on the surrounding market state. GS-LASSO effectively learns this state dependence, identifying when the execution price provides a reliable directional signal and when it should be supplemented by other publicly available information. In this sense, our approach transforms trade signing from a fixed threshold rule into a state-dependent inference problem.

The substantial improvement in accuracy matters for empirical objects used in option-market research. We first examine count-based order imbalance at the underlying-day level. For the top 20 underlying tickers by out-of-sample option volume, midpoint-implied order imbalance has an average daily cross-sectional correlation of -0.214 with true order imbalance, indicating that midpoint signing can invert the measured demand-pressure signal. By contrast, GS-LASSO-implied order imbalance has a correlation of 0.883 with true order imbalance, and the mean squared error falls by 88.0%. The improvement remains large in broader universes: for the top 100 underlying tickers, the correlation rises from 0.225 under midpoint signing to 0.838 under GS-LASSO, and the mean squared error falls by 75.2%. We also examine effective half-spreads. In the February–December out-of-sample SLAN sample, midpoint-implied relative effective half-spreads average 3.06% of the midpoint, while the true-direction value is only 0.75%. GS-LASSO-implied effective half-spreads are 1.09%, substantially closer to the true-direction benchmark.

These results show that trade-signing errors are not merely label errors at the individual-trade level; they materially change measured demand pressure and transaction costs.

We then examine whether signing errors can also affect substantive economic inference. Following Bryzgalova et al. (2023), we compute gross profits designed to capture whether SLAN order flow predicts subsequent option-price movements. Applying the Muravyev (2016) quote-based direction procedure used by BPS to our 2024 PHLX SLAN sample yields positive gross profits at every horizon, with statistically significant estimates from the intraday horizon through ten trading days. Holding fixed the trades, valuation prices, size weights, holding horizons, sample filters, and profit construction, the estimates become statistically insignificant when the inferred direction is replaced by the exchange-recorded initiator side. Thus, within our exchange-labeled sample, inference about the gross timing component of SLAN trading is sensitive to trade-direction measurement.[3]

The improvement in accuracy also allows us to revisit the economics of price improvement in option auctions. Price improvement creates a puzzle because liquidity providers frequently offer executions more favorable to the initiating trader than the prevailing quoted bid or ask, apparently giving up part of the compensation embedded in the quoted spread. We provide two pieces of evidence. First, price improvement varies with the displayed quote environment: when PHLX posts both sides of the displayed national best bid and offer (NBBO), price improvement is higher and more concentrated near the half-spread threshold. Second, despite substantial price improvement, liquidity providers earn positive short-horizon mark-to-midpoint profits in aggregate. Decomposing profits into execution and position components shows that high-PI trades are associated with positive and increasing post-trade position profits, while low-PI trades are associated with increasingly negative position outcomes. Taken together, these results suggest that price improvement is related to trade selection and position risk, rather than simply reflecting uniform competitive rent dissipation.

[3]This exercise does not re-estimate the broader national sample in Bryzgalova et al. (2023) or reassess their net-profit evidence. Its purpose is to isolate the role of direction measurement within our 2024 PHLX SLAN sample.

Our paper contributes to three strands of literature. First, it contributes to the trade-direction classification literature. Trade direction is a fundamental object in market microstructure because it underlies empirical measures of order imbalance, price impact, and trading costs (Hasbrouck, 1991; Huang and Stoll, 1997). The classic literature infers the initiator side from transaction prices, quotes, and short-horizon price movements (e.g., Lee and Ready, 1991; Ellis, Michaely, and O'Hara, 2000). In the options market, Savickas and Wilson (2003) show that quote-based rules are informative in earlier data, and a large subsequent literature uses signed option order flow to study information transmission, expected option returns, and trading costs (e.g., Coval and Shumway, 2001; Hu, 2014; Muravyev, 2016; Muravyev and Pearson, 2020). More recent evidence shows that even in equities latency and quote–trade sequencing can materially affect public trade classification (Battalio, Holden, Pierson, Shim, and Wu, 2026). We contribute to this literature by providing direct evidence, using exchange-recorded trade direction, that widely used public signing rules break down in a major modern options setting. We then recast trade classification as a conditional inference problem and show that correcting trade direction materially improves order-imbalance and transaction-cost measurement and can alter inference about the gross timing component of retail-proxy option order flow.

Second, our paper contributes to the literature on option-auction market structure. Bryzgalova et al. (2023) document that retail option trading now accounts for a substantial share of market activity and is heavily intermediated by a small number of wholesalers, and introduce SLAN orders as a proxy for retail option trading. A growing literature studies exchange-run option auctions, internalization, and competition for retail order flow, showing that auction mechanisms affect execution quality, internalization rents, allocative efficiency, and displayed quote competition (e.g., Hendershott et al., 2026; Ernst and Spatt, 2026; Ernst, Spatt, and Sun, 2025; Anand and Muravyev, 2024; Baldauf, Mollner, and Zhou, 2024). In particular, Ernst and Spatt (2026) examine price-improvement auctions and designated market maker routing, highlighting that competition in this setting remains imperfect. Our paper complements this literature by showing that understanding modern option execution requires not only analyzing how auctions affect prices

and competition, but also measuring trade direction correctly once auction mechanisms alter the informational content of execution prices. We also provide evidence that price improvement is related to displayed quote conditions, short-horizon position outcomes, and model-implied directional signals. In addition, the exchange-labeled setting allows us to isolate how trade-direction measurement affects inference about the gross timing component of SLAN order flow.

Third, our paper contributes to the literature on machine learning in finance and market microstructure. A large body of work applies machine-learning methods to asset pricing and return prediction (e.g., Light, Maslov, and Rytchkov, 2017; Chinco, Clark-Joseph, and Ye, 2019; Freyberger, Neuhierl, and Weber, 2020; Gu, Kelly, and Xiu, 2020; Kozak, Nagel, and Santosh, 2020; Avramov, Cheng, and Metzker, 2023; Chen, Pelger, and Zhu, 2024; Han, He, Rapach, and Zhou, 2024; Feng, Giglio, and Xiu, 2020; Giglio, Liao, and Xiu, 2021). More recently, machine learning has been used to recover latent microstructure objects that are difficult to observe directly (e.g., Easley, López de Prado, O'Hara, and Zhang, 2021; Bogousslavsky, Fos, and Muravyev, 2024; Bartlett and O'Hara, 2024). Closest to our setting, Bogousslavsky et al. (2024) use XGBoost to measure informed trading intensity, while Bartlett and O'Hara (2024) use machine learning to identify hidden liquidity; we build on this line of work by using an interpretable machine-learning framework to recover trade direction in auction-based option markets.

The remainder of the paper proceeds as follows. Section 2 develops the market-structure motivation and the GS-LASSO framework. Section 3 describes the data, documents midpoint-rule failure, evaluates GS-LASSO, and studies the measurement and inference consequences for order imbalance, effective spreads, and gross performance. Section 4 studies the economics of price improvement, including displayed quote venue and short-horizon mark-to-midpoint profits. Section 5 examines robustness across model specifications. Section 6 concludes.

2 Trade Direction Identification

This section first shows that modern option-market design weakens the standard link between execution prices and trade initiators, limiting the reliability of midpoint-based trade signing. We then introduce GS-LASSO, a trade-direction identification method that combines flexible nonlinear modeling with disciplined feature screening to infer the initiator side of completed trades from publicly observable market conditions around each trade.

2.1 Option Auctions, Price Improvement, and Midpoint Failure

This subsection develops the market-structure mechanism behind midpoint-rule failure. We first describe how auction-based option execution weakens the standard link between transaction prices and trade initiators, and then formalize how price improvement can make midpoint signing indeterminate or mechanically reverse the true trade direction.

2.1.1 Auction-Based Option Execution

Retail option trading has become a central component of options markets, yet standard trade-direction classification methods have not been directly validated in this setting. Bryzgalova et al. (2023) show that retail trading accounts for more than 60% of option-market volume and that this order flow is routed largely through a small group of wholesalers. The institutional setting differs sharply from equities in a way that is directly relevant for price-based trade signing. In equities, wholesalers can often internalize marketable retail orders off exchange. In listed options, by contrast, marketable retail orders are typically exposed to exchange mechanisms before execution.[4] As a result, the execution price of an option trade reflects not only the customer's trade direction but also auction competition among liquidity providers. This additional source of price formation

[4]The distinction is institutional. Retail stock orders may be executed by off-exchange market makers. Retail listed-option orders, by contrast, are generally executed on options exchanges, reflecting clearing, regulatory, and operational constraints in standardized listed options.

weakens the standard link between price location and initiator side.

Our empirical validation focuses on Single-Leg Auction Non-ISO (SLAN) trades, the Options Price Reporting Authority (OPRA) trade condition used by Bryzgalova et al. (2023) to proxy for retail option auction flow. The matched PHLX data provide the exchange-recorded initiator side for these trades, allowing us to evaluate public trade-direction inference methods directly. Section 3 describes the sample construction and economic importance of this segment.

In these auctions, liquidity providers compete through exchange-run responses rather than solely through displayed quotes.[5] Recent studies of option auctions and internalization emphasize this competition for retail order flow (Hendershott et al., 2026; Ernst and Spatt, 2026; Ernst et al., 2025). This auction competition can improve execution prices relative to the prevailing displayed quotes. For trade signing, the key implication is that such price improvement (PI) moves transaction prices toward, and sometimes across, the midpoint.

2.1.2 Price Improvement and Midpoint Failure

Let $q_i \in \{0,1\}$ denote the true initiator side, where $q_i = 1$ indicates a buyer-initiated trade and $q_i = 0$ indicates a seller-initiated trade. Let B_i and A_i denote the prevailing option best bid and ask immediately before trade i, let P_i denote the execution price, and let $S_i = A_i - B_i$ denote the quoted spread. For trades with $S_i > 0$, we measure price improvement as a fraction of the quoted spread:

$$\mathrm{PI}_i = \begin{cases} \dfrac{A_i - P_i}{S_i}, & \text{if } q_i = 1, \\ \dfrac{P_i - B_i}{S_i}, & \text{if } q_i = 0. \end{cases} \tag{1}$$

Thus, a buyer receives price improvement when the trade executes below the ask, whereas a seller receives price improvement when the trade executes above the bid.

[5] Rule-level details are provided in Online Appendix A. The key point for our analysis is that exchange-run auction responses can improve on the prevailing displayed quotes. Hendershott et al. (2026) provide additional institutional details on SLAN trades and related auction mechanisms across options exchanges.

Standard quote-based signing rules rely on a simple price-location logic. In dealer models, the bid–ask spread compensates liquidity suppliers for order-processing costs, inventory risk, and adverse selection (Garman, 1976; Amihud and Mendelson, 1980; Ho and Stoll, 1981; Glosten and Milgrom, 1985; Easley and O'Hara, 1987). This logic implies that buyer-initiated trades tend to execute on the ask side of the spread, while seller-initiated trades tend to execute on the bid side. Price improvement does not necessarily violate this mapping. If the improvement is less than one-half of the quoted spread, a buyer still executes above the midpoint and a seller still executes below the midpoint. The mapping breaks down, however, once price improvement reaches one-half of the spread.

The breakdown is mechanical. Let $M_i = (A_i + B_i)/2$ denote the quote midpoint, and let $\hat{q}_i^M$ denote the direction assigned by the midpoint rule: $\hat{q}_i^M = 1$ if $P_i > M_i$ and $\hat{q}_i^M = 0$ if $P_i < M_i$; trades at the midpoint require an auxiliary rule (Lee and Ready, 1991; Savickas and Wilson, 2003; Muravyev, 2016). From equation (1), if $q_i = 1$, then

$$P_i = A_i - \mathrm{PI}_i S_i = M_i + \left(\frac{1}{2} - \mathrm{PI}_i\right) S_i.$$

Thus, when $q_i = 1$ and $\mathrm{PI}_i > 1/2$, the trade executes below the midpoint and the midpoint rule assigns $\hat{q}_i^M = 0$. Similarly, if $q_i = 0$, then

$$P_i = B_i + \mathrm{PI}_i S_i = M_i + \left(\mathrm{PI}_i - \frac{1}{2}\right) S_i.$$

Thus, when $q_i = 0$ and $\mathrm{PI}_i > 1/2$, the trade executes above the midpoint and the midpoint rule assigns $\hat{q}_i^M = 1$. When $\mathrm{PI}_i = 1/2$, the trade executes exactly at the midpoint and price location does not identify the initiator side. Auction price improvement therefore creates a direct channel through which price-based trade-signing rules can misclassify the initiator side.

The condition above is mechanical. Section 3 shows that it is empirically first-order in PHLX SLAN trades, motivating the conditional signing framework developed next.

2.2 GS-LASSO for Trade-Direction Identification

We next develop our machine-learning framework for recovering trade direction once execution price alone is no longer sufficient. The mechanism in Section 2.1 shows that auction price improvement can make the midpoint rule indeterminate or mechanically reverse the true direction, so trade signing becomes a conditional inference problem in which the initiator side must be recovered from the public market state surrounding the trade.

2.2.1 Model Motivation

For each completed trade i, let x_i denote the vector of public market information observed around the trade, including option quotes and trades, underlying-market information, contract characteristics, and market-state variables. With q_i denoting the initiator side defined above, the objective is to recover the conditional probability

$$\Pr(q_i = 1 \mid x_i),$$

and use it to classify trade direction. Because auction outcomes depend on price improvement and competitive responses that are not fully observed in public data, this conditional-probability formulation replaces a deterministic rule based on execution price alone.

This conditional probability is likely to be nonlinear and state dependent. Auction executions aggregate the responses of participating liquidity providers, and the auction duration is only 100 milliseconds, suggesting that responses are largely automated rather than manually determined. Such automated responses are likely to depend on threshold-style rules involving spreads, depth, volatility, trade size, short-horizon price dynamics, and contract characteristics. This motivates using tree-based methods as flexible reduced-form approximations to the public component of these automated response rules.[6] XGBoost is well suited for this task because it captures nonlinearities,

[6] More details can be found in Online Appendix D.

threshold effects, and high-order interactions, while shrinkage, subsampling, and regularization help mitigate overfitting (Friedman, 2001; Chen and Guestrin, 2016). Consistent with these advantages, recent finance studies use XGBoost and related machine-learning methods to recover latent trading objects such as informed trading intensity and hidden liquidity (Bogousslavsky et al., 2024; Bartlett and O'Hara, 2024).

The challenge is that flexibility alone is not enough. The public information set contains many correlated market variables, and not all of them have stable incremental content for trade-direction recovery. A low-complexity XGBoost model with fewer boosting rounds and shallower trees is more disciplined and less prone to overfitting, but may miss nonlinearities, interactions, and rare predictive patterns. A high-complexity model with more boosting rounds and deeper trees is more flexible, but has greater scope to exploit irrelevant predictors and chance correlations. Standard tuning of boosting rounds and tree depth moves the model along this tradeoff rather than resolving it.

To address this tradeoff, we propose Guided SHAP-LASSO (GS-LASSO) to separate feature discipline from model flexibility. GS-LASSO is a general feature-screening and re-estimation framework that can be combined with a broad class of machine learning models. The name reflects the role of the first-stage model: a low-complexity machine learning model uses SHAP (SHapley Additive exPlanations; Lundberg and Lee, 2017) values to guide LASSO feature screening before a high-complexity machine learning model is estimated on the selected predictors. The next subsection describes our XGBoost-based implementation of the GS-LASSO procedure in detail.[7]

2.2.2 GS-LASSO Procedure

While GS-LASSO is designed as a general feature-screening and re-estimation framework, our empirical implementation instantiates the procedure with XGBoost. Specifically, we use gradient-boosted trees twice: first as a low-complexity representation model and then as a high-complexity

[7]The detailed mathematical formulation is provided in Online Appendix E.

final classifier. SHAP and LASSO form the intermediate screening stage. SHAP translates the first-stage nonlinear predictions into feature-level contributions, and LASSO selects the predictors whose contributions remain robust conditional on other predictors. This structure allows the final classifier to use a flexible nonlinear model without exposing it to the full noisy predictor set.

First, we estimate a low-complexity XGBoost model on the full set of candidate predictors and extract its SHAP values. The purpose of this first-stage model is not to maximize classification accuracy, but to extract stable predictive signals under a tight complexity constraint. This design is consistent with the boosting literature (Zhang and Yu, 2005; Bühlmann and Hothorn, 2007), which views early stopping and restrictions on model complexity as forms of regularization. By limiting both the number of boosting rounds and tree depth, the first-stage model has less scope to fit chance correlations or noisy patterns. Its SHAP values then provide a feature-level representation of the strongest nonlinear signals in the public information set.

Second, we apply cross-validated logistic LASSO to the SHAP values. The first-stage XGBoost model provides a disciplined nonlinear representation, but it does not impose sparsity at the feature level. Even a regularized tree ensemble may assign nonzero contributions to weak or spuriously correlated predictors. The SHAP representation makes feature screening tractable by placing feature contributions on a common additive scale. The LASSO step then retains features whose first-stage predictive contributions remain robust conditional on other features and removes variables whose apparent contributions are not stable.

Third, we re-estimate a high-complexity XGBoost classifier using the selected raw predictors. This step is important because SHAP-LASSO is used for feature selection rather than final prediction. The final classifier is not constrained to be linear in SHAP values. Once noisy or spurious predictors have been screened out, the second-stage XGBoost model can use more boosting rounds and deeper trees to learn nonlinearities, high-order interactions, and rare predictive patterns among the selected predictors. GS-LASSO therefore separates feature discipline from model flexibility: the first two steps restrict the predictor space, while the final step increases

complexity only within that selected space.

This procedure is conceptually related to the finance literature on model complexity. Although our objective is trade-direction identification rather than return prediction, our design is motivated by a similar tradeoff between flexibility and disciplined regularization. Kelly, Malamud, and Zhou (2024) emphasize that flexible models can improve predictive performance when combined with shrinkage, while Kelly and Malamud (2025) distinguish nominal complexity from effective complexity. GS-LASSO brings this logic to a measurement setting: SHAP-guided LASSO screening restricts the effective predictor space, and the final XGBoost classifier uses nonlinear flexibility only among the selected features.

Online Appendix F provides simulation evidence supporting this screening logic. In a setting with nonlinear signals, rare predictive patterns, noisy predictors, and unstable false features, GS-LASSO substantially reduces the number of irrelevant variables retained by the final model and significantly improves both out-of-sample overall accuracy and rare-pattern accuracy after high-complexity re-estimation.

2.2.3 Implementation and Evaluation

In the main empirical implementation, we estimate GS-LASSO separately within five observable execution-price-location groups. The partition is determined only by the transaction price relative to the contemporaneous bid, ask, and midpoint. Specifically, trades are classified as executing at the bid, at the ask, exactly at the midpoint, inside the spread below the midpoint, or inside the spread above the midpoint. These groups represent distinct signing problems. At the bid or ask, price location provides a strong but potentially reversible signal; at the midpoint, price location is indeterminate; and inside the spread, the informativeness of price location depends on the extent of price improvement. Estimating separate models across these groups allows the selected features and nonlinear interactions to differ across observable price-location states, rather than forcing one global mapping from public information to trade direction. Importantly, the partition uses only

public price and quote information and does not depend on the true trade direction or realized price improvement.

We implement GS-LASSO in a rolling-window design. Let t index non-overlapping calendar periods. To infer trade directions in period $t+1$, we use observations from period t as the training sample. Specifically, let $\mathscr{W}_t$ denote the set of trades in period t. All model estimation, SHAP decomposition, and LASSO feature selection are performed using only observations in $\mathscr{W}_t$. The resulting selected predictor set and fitted classifier are then applied to trades in $\mathscr{W}_{t+1}$. This rolling design allows the mapping from public market information to trade direction to evolve over time while preserving a strict separation between model construction and evaluation.

We evaluate performance using out-of-sample classification accuracy:

$$\text{Accuracy} = \frac{1}{N_{\mathscr{W}_{t+1}}} \sum_{i \in \mathscr{W}_{t+1}} \mathbf{1}\{\hat{q}_i = q_i\}, \tag{2}$$

where q_i is the exchange-recorded initiator side, $\hat{q}_i$ is the predicted initiator side, and $N_{\mathscr{W}_{t+1}}$ is the number of trades in the evaluation period.

This design ensures that no observation in $\mathscr{W}_{t+1}$ is used to choose features, estimate trees, or otherwise construct the classifier. Reported performance therefore reflects genuine out-of-sample identification. We benchmark GS-LASSO against midpoint-based and exchange-quote-based signing rules on the same set of trades to test whether a richer public information set can recover the exchange-recorded trade direction more accurately in an auction environment where execution price alone is insufficient.[8]

[8] We also examine alternative implementations, including static estimation windows and a plain XGBoost classifier without SHAP-LASSO screening. These exercises distinguish gains from the rolling design, from nonlinear modeling alone, and from the combination of nonlinear representation and structured feature selection.

3 Data and Empirical Results

This section brings the mechanism in Section 2 to the data. We first describe our dataset and the public market information used to construct our trade-direction classifier. We then use the exchange-recorded initiator side to document the failure of midpoint-based signing in auction-based option execution. Next, we evaluate the main GS-LASSO specification and examine the public signals that contribute to its predictions. Finally, we show that improved trade-direction measurement materially affects signed order imbalance, effective spreads, and inference about the gross timing component of SLAN trading.

3.1 SLAN Auction Sample

Our analysis combines auction-level messages from Nasdaq PHLX LLC (PHLX) with transaction and quote data from the Options Price Reporting Authority (OPRA) for January 1 to December 31, 2024.[9] PHLX is among the largest U.S. options exchanges by equity-options trading volume and is an economically important venue for listed option trading. The PHLX feed records nanosecond-level information on orders added, modified, canceled, filled, and auctioned across all tradable option series. We extract all simple-order auction messages, including order direction, option contract characteristics, timestamps, auction prices, and auction volumes, and match executed auction orders to OPRA transaction records using execution timestamps and option identifiers.

We focus on Single-Leg Auction Non-ISO (SLAN) trades, identified using the SLAN trade condition in OPRA, following Bryzgalova et al. (2023). This segment is central to our analysis because the matched PHLX data reveal the exchange-recorded true direction of the initiating customer order, providing the ground-truth needed to evaluate trade-classification methods. The raw sample contains 49.71 million SLAN trades executed on PHLX in 2024. SLAN trades account for 63.10% of total option volume on PHLX during our sample period, closely matching the

[9] We obtain the data from ThetaData and SpiderRock, two authorized OPRA data providers.

estimate in Bryzgalova et al. (2023) that SLAN represented over 60% of total market volume in early 2020. This similarity indicates that the SLAN environment is not a narrow corner of the market, but a representative segment of modern listed-option retail execution.

We further match the SLAN sample to OPRA quote data, underlying trade and quote data from TAQ, and additional SpiderRock option-market variables. These data allow us to construct 445 public market-state variables capturing contract and trade characteristics, option quote and depth conditions, short-horizon option and underlying price dynamics, aggregate market conditions, and vendor-based fair-value, volatility-surface, and quote-state measures.[10] To ensure data quality, we exclude trades executed outside the quoted spread, observations with non-positive spreads or prices, and trades with non-positive sizes or sizes above 250,000 contracts.[11] These filters remove only about 0.93% of observations. After restricting attention to stock and exchange-traded fund (ETF) options and applying the filters, the final sample contains 46.68 million SLAN trades.

Table 1 reports summary statistics for the final sample. The dataset contains an average daily trading volume of 1.57 million contracts, with substantial day-to-day variation. Trading is concentrated in at-the-money contracts, which account for 61.54% of total volume, while out-of-the-money and in-the-money options account for 24.55% and 13.88%, respectively. Calls dominate the sample, representing about 65% of both volume and trade count, and buy and sell order volume is nearly balanced, with buyer-initiated trades accounting for 50.61% of daily contract volume on average.

Stock and ETF options jointly account for 94.64% of SLAN activity on PHLX. Within the final stock-and-ETF sample reported in Table 1, stock options account for 72.60% of contract volume, while ETF options account for the remaining 27.40%. Although stock options account for roughly 80% of daily trades, ETF option trades are larger on average, implying a higher mean trade size for ETFs than for individual stocks. Across the combined stock and ETF sample, the average trade size is 8.49 contracts, and 45.64% of transactions occur at the minimum size of one contract,

[10] Detailed feature definitions are provided in Online Appendix M.

[11] The limit of 250,000 contracts follows Nasdaq PHLX position rules; see Section 13(g) of the PHLX rulebook.

consistent with the retail character of SLAN trading. Finally, ETF options exhibit tighter quoted spreads than stock options.

Overall, the SLAN dataset provides a large-scale and economically representative setting for studying trade classification in auction-based options markets. Its combination of exchange-recorded true direction and rich public market information makes it well suited to evaluating both midpoint-based rules and data-driven alternatives.

3.2 Midpoint Failure in Auction-Based Execution

This subsection brings the midpoint-failure mechanism in Section 2.1 to the data. Using the exchange-recorded initiator side, we first evaluate standard public trade-signing rules and show that their accuracy is only modestly above chance. We then show why: price improvement places a large mass of SLAN trades at or beyond the one-half-spread threshold, where midpoint signing either becomes indeterminate or mechanically reverses. Finally, we show that this failure varies across observable price-location and market states, motivating the five execution-price-location groups and the rolling conditional classifier evaluated in the next subsection.

3.2.1 Benchmark Failure and the Price-Improvement Mechanism

Table 2 reports the accuracy of standard public trade-signing rules relative to the exchange-recorded true direction. Panel A shows that the midpoint rule performs poorly in the SLAN setting. Its average daily accuracy is only 55.34%, only modestly above chance, and the failure is persistent over time: the interquartile range is narrow, from 53.85% to 57.03%, and even the 95th percentile of daily accuracy is only 59.05%. The poor performance is also broad-based across trade types. Accuracy is 54.75% for buy trades, 55.95% for sell trades, 55.67% for calls, and 54.60% for puts. Thus, midpoint signing is not failing because of a few unusual days or a particular subset of trades; it performs poorly throughout the sample and across economically important trade categories.

Panel B of Table 2 shows that the problem is not solved by a standard quote refinement. The exchange-quote refinement of Muravyev (2016) improves average daily accuracy from 55.34% to 60.16%, indicating that exchange-specific quote information contains additional signing content.[12] However, the improvement is limited: the rule still misclassifies roughly 40% of trades on average, and its accuracy remains close to 60% across buys, sells, calls, and puts. These results show that the failure of public signing rules is not simply a consequence of using the weakest possible price-based benchmark. Even a refined quote-based method remains far from sufficient in auction-based option executions.[13]

The price-improvement mechanism explains why these public signing rules perform so poorly. Section 2.1 shows that midpoint signing becomes indeterminate when price improvement equals one-half of the quoted spread and mechanically reverses the true direction when price improvement exceeds one-half. Panel A of Figure 1 shows that this threshold region contains a large share of SLAN trades. In the full sample, 13.86% of trades receive exactly 50% price improvement, so the transaction price is at the quote midpoint and price location no longer identifies the initiator side. Another 30.67% of trades receive more than 50% price improvement, so the transaction price crosses the midpoint and midpoint signing mechanically reverses the true initiator side. Thus, 44.53% of SLAN trades lie in the region where midpoint signing either loses its directional signal or reverses it.

Panel A of Table 3 shows that large price improvement is a central feature of the SLAN environment, not an isolated tail event. Average price improvement is 43.23% of the quoted spread in the full sample; the median is 40.00%, the 75th percentile is 60.00%, and the 95th percentile is 100.00%. These summary statistics line up directly with the CDF evidence: a large mass of

[12] Grauer, Schuster, and Uhrig-Homburg (2026) propose an additional structure-based procedure for classifying option demand. When we implement their published rules in our 2024 PHLX SLAN sample and evaluate the resulting signs against exchange-recorded direction, the procedure achieves approximately 62% accuracy. Because their reported results use a different sample and empirical setting, we interpret this exercise as evidence on the procedure's performance in our auction sample rather than as a direct replication comparison. Online Appendix K provides implementation details.

[13] Online Appendix G shows that the low accuracy of midpoint signing is not driven by quote mismatching or timestamp misalignment, despite the importance of such frictions for public trade classification in other settings (Battalio et al., 2026).

trades receives price improvement near or beyond the midpoint-failure threshold. The pattern is also similar across initiator side and option type. Average price improvement is 43.46% for buys, 43.00% for sells, 43.07% for calls, and 43.53% for puts, and each group has a 75th percentile of 60.00% and a 95th percentile of 100.00%. Panel B of Figure 1 provides the corresponding distributional evidence, showing that the CDFs for calls, puts, buys, and sells are close to one another. Together, Table 2, Figure 1, and Table 3 show that the failure of public signing rules is mechanically tied to the prevalence of large price improvement and is broad-based across the SLAN sample.

3.2.2 Price-Location States and Benchmark Refinements

Panel B of Table 3 connects the midpoint-failure mechanism to observable execution-price locations. Relative to the contemporaneous NBBO, trades fall into five public price-location states: at the ask, at the bid, exactly at the midpoint, inside the spread below the midpoint, or inside the spread above the midpoint. These states are not arbitrary bins; they correspond to distinct signing problems because the information content of price location differs across them. At the bid or ask, price location provides a strong directional signal, but that signal can be reversed by full price improvement: 38.86% of at-ask trades and 40.77% of at-bid trades receive full price improvement, so the midpoint-implied direction is reversed in a large fraction of endpoint executions. Exact-midpoint trades, which account for 6.47 million observations, are mechanically associated with 50% price improvement and therefore cannot be signed from price location alone. For inside-spread trades away from the midpoint, price location remains informative but only conditionally: average price improvement is 43.19% below the midpoint and 43.54% above the midpoint, and 32.82% and 33.99% of these trades, respectively, receive price improvement above 50%.

These price-location patterns also explain the limited improvement from the exchange-quote refinement in Muravyev (2016). Among trades executed at the NBBO midpoint, the exchange quote is also at its own midpoint in 44.82% of cases, leaving the refinement without a local price-

location signal. Its unconditional accuracy on exact-midpoint trades is only 34.69%. Conditional on a nonzero exchange-quote sign, accuracy rises to 63.03%, but this conditioning applies to only about 55% of exact-midpoint trades. Thus, quote location is useful but incomplete. This evidence motivates the five execution-price-location groups used in our main implementation: endpoint trades, where the price signal is strong but reversible; midpoint trades, where the price signal is indeterminate; and inside-spread trades, where the price signal depends on the extent of price improvement. Estimating the classifier separately across these observable states allows the selected features and nonlinear interactions to differ across the distinct signing problems created by auction price improvement.

3.2.3 State Dependence in Midpoint Failure

Midpoint failure is also state dependent beyond price location itself. Panel C of Figure 1 shows that ETF options receive more aggressive price improvement than stock options: their average daily price improvement is 44.38% of the quoted spread, compared with 43.12% for stock options. Similarly, 47.79% of ETF-option trades receive at least 50% price improvement, compared with 44.00% for stock-option trades. The level of price improvement differs across the two asset classes, but the signing implication is common to both: trades at or beyond the midpoint cutoff are frequent in each group.

Time variation provides another dimension of state dependence. Figure 2 shows that midpoint accuracy is highest during the most active part of the session, especially shortly after the market opens, and declines afterward. The stock–ETF decomposition in Panel B shows that this intraday pattern is not purely an aggregate effect: both groups exhibit declining accuracy, but stock options remain consistently easier to sign than ETF options. Calendar patterns also indicate that the price-improvement environment is not stable over time. Panel A of Figure 3 shows that price improvement is highest at the start of the trading week and lowest at the end: Monday has an average price improvement of 43.95%, compared with 42.84% on Friday. In paired week-level

tests comparing Monday and Friday within the same Monday–Friday trading week, the Monday–Friday difference is 1.05%, with a t-statistic of 4.37. Panel B of Figure 3 shows that average price improvement is relatively flat in the first half of the month and becomes higher and more variable in the second half. These patterns support a rolling-window design, because the mapping from public market conditions to trade direction is likely to evolve over time.[14]

Finally, displayed quote venue provides a cross-sectional state tied directly to the auction environment. All trades in our sample execute in PHLX auctions, but the displayed NBBO may or may not be posted on PHLX. Panel C of Table 3 and Panel D of Figure 1 show that price improvement differs across these quote-side states. When both NBBO sides are posted on PHLX, mean price improvement is 46.14% and the median is 46.67%. When neither the NBBO bid nor the NBBO ask is posted on PHLX, the corresponding values are 43.06% and 40.00%. The one-sided PHLX states lie between these cases. Thus, the probability of reaching the midpoint-failure region depends not only on contract and time-state variables, but also on the quote-side venue environment. We return to the economic interpretation of this venue pattern in Section 4.2.

Taken together, the evidence shows that midpoint-rule failure is a structural feature of the SLAN environment. Standard public signing rules perform poorly, and the reason is direct: a large share of SLAN trades lies at or beyond the 50% price-improvement cutoff where midpoint signing ties or reverses. At the same time, transaction price remains informative in a state-dependent way. Its content varies across execution-price locations, asset classes, intraday conditions, calendar time, and quote-side venue states. This evidence motivates the conditional classification approach evaluated next, which uses richer public market information to recover the initiator side of completed auction trades.

[14] Online Appendix B discusses a routing-accountability interpretation based on calendar-based execution-quality monitoring.

3.3 GS-LASSO Performance

The evidence above shows that midpoint signing fails in the SLAN setting because auction price improvement frequently weakens or reverses the link between transaction price and trade direction. We now evaluate whether our GS-LASSO specification can recover the exchange-recorded initiator side from the public market information surrounding each trade.

We implement the GS-LASSO specification using the five execution-price-location groups described above and the rolling procedure in Section 2.2. The model is estimated separately across the five observable groups and evaluated in a calendar-month rolling design. To classify trades in month $m+1$, all feature selection, tuning, and model estimation use only trades from month m. The out-of-sample period therefore begins on February 1, 2024, because January 2024 is used as the first training month, and ends on December 31, 2024. The candidate set contains 445 public variables constructed from PHLX, OPRA, TAQ, and SpiderRock data, including contract and trade characteristics, option quote and depth variables, short-horizon option and underlying price dynamics, aggregate market-state variables, and SpiderRock fair-value, volatility-surface, and quote-state measures.[15] The candidate set is restricted to public variables; variables that require the exchange-recorded true direction to construct, such as realized price improvement or indicators based on the price-improvement cutoff, are excluded.

Table 4 reports summary statistics for out-of-sample GS-LASSO accuracy. The model achieves an average out-of-sample accuracy of 80.46% across all trades. Accuracy is also balanced across trade types: the model achieves 80.14% accuracy for buy trades, 80.75% for sell trades, 80.20% for calls, and 80.94% for puts. Compared with the midpoint benchmark documented in Section 3.2, this represents an improvement of roughly 25 percentage points. The improvement is therefore not driven by a specific trade direction or option type, but is broad-based across the SLAN sample.

[15] For researchers seeking a directly implementable alternative, Online Appendix L reports five depth-3 decision trees, one for each execution-price-location group. Using ten public variables available by the transaction time, the five trees achieve a combined mean daily out-of-sample classification accuracy of 74.26%. The decision-tree rules are reported directly, and the code and fitted model files for the full 445-candidate-variable implementation are posted on the authors' websites.

Figure 4 plots daily accuracy for GS-LASSO and the midpoint rule. The GS-LASSO series remains close to 80% throughout the out-of-sample period, while midpoint signing stays near the low-accuracy region documented earlier. Model accuracy remains close to 80% across all eleven out-of-sample months, with monthly accuracy ranging from 79.17% to 81.50%. The stability of the model's performance is important because a high-dimensional classifier could, in principle, perform well on average by fitting a few periods and failing in others. That is not what we observe: the model's improvement persists across changing trading conditions, including the calendar variation in price improvement documented above.

To understand where the model improves most, Table 5 reports accuracy across subsamples defined by key contract and trading characteristics. GS-LASSO outperforms the midpoint rule in every subsample, and all differences are statistically significant at the 1% level. The gains are largest precisely in the environments where midpoint signing is most fragile. For example, for below-median option prices, GS-LASSO achieves 84.25% accuracy, compared with 51.87% for the midpoint rule, a gain of 32.38 percentage points. For below-median dollar quoted spreads, the model reaches 85.00% accuracy, compared with 54.74% for midpoint signing, a gain of 30.26 percentage points.

These patterns line up closely with the midpoint-failure mechanism. When option prices are low or quoted spreads are tight, even modest dollar price improvement can move the execution price to or across the midpoint. In those states, execution price becomes a weak sufficient statistic for trade direction, and a one-threshold signing rule performs poorly. GS-LASSO gains precisely in these settings because it conditions on the broader public market state rather than using price location alone.

The model also performs well outside the most difficult states. For above-median option prices, GS-LASSO improves accuracy by 17.81 percentage points relative to midpoint signing; for above-median quoted spreads, the gain is 19.43 percentage points. Across moneyness and bid–ask depth-ratio splits, the gains are close to 25 percentage points on both sides of the median. These results

show that the model is not exploiting a narrow feature of a particular subset of trades. Instead, it learns a stable conditional mapping from public market information to trade direction across a broad range of option contracts and liquidity environments.

Taken together, the results show that a substantial component of trade direction remains recoverable from public information even when execution price loses its traditional signing content. The gains are largest where the midpoint rule fails most, which supports the interpretation that GS-LASSO is not replacing price information with an unrelated black-box pattern. Rather, it recovers trade direction by conditioning price-location information on the market state in which the auction occurs.

3.4 Feature Contributions and Conditional Signing

We next examine which public variables are most informative for identifying trade direction. The purpose is to understand whether the model's gains are linked to economically meaningful market-state variables rather than diffuse high-dimensional patterns. We use SHAP attributions from the first-stage nonlinear model to summarize which variables contribute to the predicted probability of buyer initiation. The SHAP importance measures are based on mean absolute SHAP values and therefore capture average predictive importance rather than causal effects or directional coefficients.

Figure 5 reports the ten largest annual mean absolute SHAP values from the first-stage models used in GS-LASSO. The leading variables point to several economically interpretable dimensions of the signing problem. Quoted spreads and displayed-depth variables capture the local liquidity environment in which auction prices are formed. Trade size captures order-level exposure. Short-horizon option returns capture recent option-price dynamics around the auction. Underlying-market pressure variables capture movements in the underlying asset, while SpiderRock fair-value and volatility-surface variables capture deviations between transaction prices, quotes, and model-implied values. Taken together, these variables show that trade direction is not recovered from

execution price alone. The model combines price-location information with liquidity, depth, short-horizon price dynamics, and fair-value signals.

The strongest features in Figure 5 are closely related to the midpoint-failure mechanism in Section 2.1. When price improvement is large relative to the quoted spread, the mapping from execution price to trade direction becomes unstable. The model therefore learns that the signing content of price location depends on the local spread and liquidity state. Other highly ranked variables, such as trade size, displayed NBBO depth ratio, short-horizon option returns, and fair-value gaps, further indicate that the model conditions price information on the broader market state rather than applying a fixed price-based threshold.

Figure 6 examines whether these signals are stable. Panel A ranks the same ten features within each monthly rolling training window. The top signals are persistent throughout the year, indicating that the model relies on stable economic signal categories rather than transitory noise. At the same time, the rankings are not identical across months, which is consistent with the calendar-month rolling design. The model relies on a stable set of economic signals, but allows their relative importance to adjust as market conditions change.

Panel B of Figure 6 shows that the same variables do not play the same role across the five execution-price-location groups. For endpoint trades at the ask or bid, quoted spreads and trade size are among the most important predictors, consistent with the idea that endpoint executions carry a strong price-location signal but may still be reversed by full price improvement. For exact-midpoint trades, price location is uninformative by construction, and the model relies more heavily on displayed-depth imbalance, short-horizon option returns, underlying-market pressure, and fair-value deviations. For inside-spread trades below or above the midpoint, price position within the NBBO becomes especially important, but it is combined with option returns, depth, and fair-value variables. These patterns support the five execution-price-location group implementation: different public price-location states represent different signing problems, so the relevant predictors and nonlinear interactions differ across states.

Overall, the feature evidence reinforces the interpretation that GS-LASSO acts as a conditional signing framework rather than an unrelated black box. The model does not discard price information. Instead, it learns when price-location information should be trusted and when it should be adjusted using liquidity, depth, short-horizon price dynamics, fair-value deviations, and underlying-market information. This is precisely the setting in which traditional midpoint signing fails: once auction price improvement weakens the direct mapping from execution price to initiator side, trade direction must be inferred from the broader public market state surrounding the auction.

3.5 Measurement and Inference Consequences of Trade-Signing Errors

Trade-direction classification matters because signed trades are inputs into empirical measures of order imbalance, demand pressure, transaction costs, liquidity provision, and trading performance. The key question is therefore not only whether GS-LASSO classifies individual trades more accurately, but whether direction errors distort the empirical objects and economic conclusions constructed from signed trades. We study this question in three applications: signed order imbalance, effective spreads, and gross-performance inference. The first two compare midpoint-implied, GS-LASSO-implied, and exchange-recorded direction. The third uses a common-sample design to isolate the effect of replacing a public quote-based direction measure with the exchange-recorded initiator side.

Throughout the first two applications, let $D_i^k \in \{-1, 1\}$ denote the signed trade direction under direction input k, with $D_i^k = 1$ for buyer initiation and $D_i^k = -1$ for seller initiation.[16] We consider $k \in \{\text{True}, \text{GS}, \text{Mid}\}$, corresponding to the exchange-recorded true direction, the GS-LASSO-implied direction, and the midpoint-implied direction.

[16] Earlier sections use the binary indicator $q_i \in \{0, 1\}$, which is natural for classification and predicted probabilities. In the first two applications below, we use the conventional signed representation $D_i^k \in \{-1, 1\}$ because order imbalance and effective spreads are signed quantities.

3.5.1 Signed Order Imbalance

Signed option order flow is commonly summarized as order imbalance or demand pressure and then related to option prices, expected returns, retail trading, and market-level predictability (Muravyev, 2016; Bryzgalova et al., 2023; Ni, Pearson, Poteshman, and White, 2021). Because order imbalance is a signed sum, systematic sign errors need not average out. Instead, they can change the measured demand-pressure variable itself.

For underlying ticker r on trading day t, we define count-based order imbalance as

$$OI_{r,t}^{count,k} = \frac{1}{N_{r,t}} \sum_{i \in (r,t)} D_i^k,$$

where $N_{r,t}$ is the number of SLAN trades written on underlying ticker r on day t. Thus, each observation is an underlying-day imbalance that aggregates signed trades across all option contracts written on the same underlying stock or ETF.

Table 6 compares midpoint- and GS-LASSO-implied order imbalance with the true-direction benchmark. Underlying ticker groups are ranked by total out-of-sample option volume, and the table reports each group's share of total out-of-sample volume. The top 20 underlying tickers account for 56.46% of out-of-sample option volume, while the top 100 account for 77.45%. These groups therefore cover the most economically important part of the SLAN sample.

Midpoint-implied direction can severely distort the signed-flow signal. In the top-20 underlying-ticker group, midpoint-signed count order imbalance has an average daily cross-sectional correlation of -0.214 with true order imbalance. In other words, for the most actively traded underlyings, midpoint signing does not merely add noise; it can invert the cross-sectional demand-pressure signal. By contrast, GS-LASSO-implied order imbalance has a correlation of 0.883 with true order imbalance. Panel B shows the same pattern using mean squared error: for the top-20 group, the MSE falls from 0.0229 under midpoint signing to 0.0028 under GS-LASSO, an 88.0% reduction.

The improvement remains large in broader universes. For the top-100 underlying tickers, the

correlation with true order imbalance rises from 0.225 under midpoint signing to 0.838 under GS-LASSO, and the MSE falls by 75.2%. For the top-500 group, which accounts for 94.49% of out-of-sample volume, the correlation rises from 0.366 to 0.760, and the MSE falls by 59.3%. These results show that trade-signing errors materially affect the signed-flow variables used to measure option demand pressure, while GS-LASSO recovers a large fraction of the true order-imbalance signal.

3.5.2 Effective Spreads

Effective spreads provide a second application because transaction-cost measures also require trade direction. Let P_i denote the transaction price and M_i the prevailing quote midpoint. For any direction input k, we define the effective half-spread as

$$ES_i^k = D_i^k(P_i - M_i), \tag{3}$$

where $D_i^k \in \{-1, 1\}$ is the signed trade direction under method k.[17] This definition makes the measurement problem under midpoint signing immediate. Away from the midpoint, midpoint signing sets $D_i^{\text{Mid}} = \text{sign}(P_i - M_i)$, so

$$ES_i^{\text{Mid}} = |P_i - M_i| \geq 0.$$

This is appropriate only when price location correctly identifies the initiating, liquidity-demanding side. In price-improving auctions, however, buyer-initiated trades can execute below the midpoint and seller-initiated trades can execute above the midpoint. For these cross-midpoint trades, the true-direction effective spread is negative, reflecting an execution favorable to the initiating trader relative to the midpoint, but midpoint signing records the same trade as a positive trading cost. Thus, midpoint signing mechanically converts cross-midpoint price improvement into measured

[17] The conventional effective spread is simply $2ES_i^k$; because our interest is in comparing alternative direction inputs rather than scaling conventions, we report the effective half-spread throughout.

effective spread and biases transaction-cost estimates upward.

Table 7 reports ES_i^k under the three direction inputs. We report the measure in two units: $100 \times ES_i^k/M_i$, expressed as a percentage of the midpoint, and $100 \times ES_i^k$, expressed in cents of option premium. Multiplying by two would give the conventional full-spread version and does not affect the comparison across direction inputs.

The magnitude of the bias is large. In the February–December out-of-sample SLAN sample, the midpoint-implied relative effective half-spread is 3.06% of the midpoint, while the true-direction value is only 0.75%. The GS-LASSO-implied value is 1.09%, much closer to the true-direction benchmark. The cents measure gives the same conclusion: the midpoint-implied effective half-spread is 3.11 cents, compared with 0.83 cents using the exchange-recorded true direction and 1.32 cents using the GS-LASSO-implied direction.

Thus, midpoint signing substantially overstates transaction costs in auction-based option trades, producing an effective-spread estimate 4.08 ($= 3.06\%/0.75\%$) times as large as that based on the exchange-recorded true trade direction. This approximately fourfold overstatement closely matches the magnitude documented by Barardehi et al. (2025) for retail option trades with limit-order usage.[18] Against this benchmark, GS-LASSO does not fully eliminate the measurement error induced by trade signing, but substantially narrows the gap between the estimated and true-direction effective spreads. Together with the order-imbalance evidence, these results show that signing errors materially distort empirical quantities constructed from option trades. We next examine whether the choice of direction measure can also alter inference about gross performance.

3.5.3 Trade-Direction Measurement and Gross-Performance Inference

Trade direction also enters the construction of gross dollar performance for retail-proxy option order flow. Following Bryzgalova et al. (2023, hereafter BPS), gross performance is based on

[18] Using the estimates reported in Table 3 of Barardehi et al. (2025), the ratio of the conventional effective-spread estimate to the unadjusted true-direction estimate is $2.59/0.60 = 4.32$.

the signed change from the option midpoint at the time of the trade to its terminal value at the end of the holding horizon. Because the initial value is the quote midpoint rather than the actual transaction price, this measure abstracts from the opening execution-cost component and captures gross performance before those transaction costs.

We apply the BPS construction to our February–December 2024 PHLX SLAN sample. Within each holding horizon, the common sample consists of trades to which the two-stage quote procedure of Muravyev (2016), used by BPS in their gross-performance analysis, assigns a direction. The procedure first signs trades relative to the contemporaneous NBBO midpoint and, for trades at that midpoint, applies the quote rule relative to the midpoint of the executing-exchange BBO. Trades that remain unsigned after the second stage are excluded. We hold fixed the trades, capped trade sizes, the initial midpoint selected by that procedure, terminal valuation prices, sample filters, and daily aggregation procedure; only the direction input changes. Table 8 compares Muravyev (2016) Quote-Based Direction with the exchange-recorded initiator side.

Using Muravyev (2016) Quote-Based Direction, mean daily gross performance is positive at every horizon. The estimates range from \$2.385 million intraday to \$3.059 million at ten trading days and are statistically significant at the 1% level through the ten-day horizon. The expiration estimate is also positive, at \$1.397 million per day, but is not statistically significant. Thus, within our PHLX sample, the Muravyev (2016) procedure reproduces the qualitative pattern of positive gross performance documented by BPS.

The conclusion changes when the same trades are signed using the exchange-recorded initiator side. The corresponding estimates lie between -0.276 and 0.236 million dollars per day through the ten-day horizon and equal -0.215 million dollars at expiration; none is statistically significant. Because every other component of the calculation is held fixed, the comparison isolates the role of direction measurement. This exercise does not reassess BPS's use of auction trades as a proxy for retail option activity, their broader sample, or their net-performance evidence. Rather, it shows that, in our 2024 PHLX auction sample, inference about positive gross performance is sensitive to

the trade-direction measure used.[19]

4 The Economics of Price Improvement in Option Auctions

Section 3 shows that price improvement is pervasive in modern option auctions and that accurate trade-direction measurement is essential for studying auction-based execution. In this section, we use the exchange-recorded true direction to study the economics of PI. We ask two related questions. First, how is PI related to the displayed quote environment in which the auction occurs? Second, can large execution concessions coexist with positive short-horizon trading outcomes for liquidity providers?

4.1 The Economic Puzzle of PI

PI is a pervasive feature of modern option auctions. As shown in Section 3, a substantial fraction of trades execute at prices strictly inside the quoted spread, and many receive improvements exceeding one-half of the spread. These patterns indicate that liquidity providers frequently transact at prices that are more favorable to the initiating trader than the prevailing quoted bid or ask.

This evidence raises a natural economic puzzle. In standard market microstructure models, liquidity providers earn the bid–ask spread as compensation for order-processing costs, inventory risk, and adverse selection (Garman, 1976; Amihud and Mendelson, 1980; Ho and Stoll, 1981; Glosten and Milgrom, 1985; Easley and O'Hara, 1987). Under this view, execution prices are expected to remain close to the bid or ask, so that the quoted spread serves as a primary source of protection against trading risks. Systematic PI appears difficult to reconcile with this logic because it implies that liquidity providers frequently offer execution concessions relative to the prevailing

[19] For completeness, Online Appendix J reports corresponding estimates based on the baseline GS-LASSO-implied direction and the fixed public-data decision trees. Because the baseline GS-LASSO classifier is designed for ex-post trade signing and may use short-horizon post-trade public signals, we treat its estimates as a descriptive downstream application rather than as an independent test of subsequent option-price predictability.

quoted terms, thereby giving up part of the compensation embedded in the quoted spread.

The puzzle is therefore not simply that PI occurs, but that it is both frequent and substantial. If wholesalers cannot recover these concessions through other channels, then aggressive PI should compress profits and be difficult to sustain in equilibrium. One possibility is that PI simply reflects competitive pressure for order flow. Under this view, wholesalers bid away rents in order to win execution, and PI represents a transfer to investors with little offsetting benefit to liquidity providers.

An alternative possibility is that PI is used selectively. Rather than applying a uniform pricing rule across all incoming orders, wholesalers may condition the amount of PI on information about the trade and the surrounding market state. Under this view, large concessions relative to the quoted bid or ask need not imply weak pricing or indiscriminate competition. Instead, PI may be part of a broader strategy through which wholesalers manage trading risk, allocate liquidity, and balance execution concessions against subsequent outcomes on the positions they acquire.

4.2 Displayed Quote Venue and Price Improvement

The quote-side venue evidence provides a first indication that PI is linked to the competitive state surrounding the auction. All trades in our sample execute on PHLX, but the displayed national best bid and offer need not be posted on PHLX. This distinction is useful because it separates the venue running the auction from the venue or venues setting the displayed inside market. If auction competition is related to the displayed quote environment, then PI should vary with whether PHLX is also posting the national best bid and/or offer.

This intuition is closely related to the internalization and auction literature. Anand and Muravyev (2024), for example, show that internalization can affect quote competition. In listed options, where retail orders ultimately execute on exchanges, price improvement auctions provide another channel through which competition for retail flow can occur: rather than competing only through displayed quotes, liquidity providers can compete inside the auction response process. The

venue composition of the displayed best bid and offer therefore captures a public quote-side state that may be related to the amount and form of PI.[20]

Panel C of Table 3 and Panel D of Figure 1 show that PI differs meaningfully across quote-side venue states. When both displayed NBBO sides are posted on PHLX, mean PI is 46.14% of the quoted spread and the median is 46.67%. When neither the displayed bid nor the displayed ask is posted on PHLX, the corresponding values are 43.06% and 40.00%. The one-sided PHLX states lie between these cases. The distributional evidence also differs across states: the PHLX-both state has less mass in the low-PI region and is more concentrated near the half-spread cutoff, while the no-PHLX state has more mass near zero and in the far right tail.

These patterns suggest that the displayed quote venue captures more than a mechanical percentage concession. It identifies a quote-side liquidity state in which the auction venue's relation to the displayed inside market differs. When PHLX is also posting both displayed quote sides, the auction takes place on the same venue that is setting the inside market, and PI is more concentrated around intermediate improvements. When the displayed inside market is away from PHLX, the auction more often interacts with another venue's displayed quote, producing a more polarized PI distribution. This evidence is descriptive, but it supports the view that auction-based PI is linked to displayed quote competition and venue-specific liquidity conditions. It also helps explain why midpoint signing fails heterogeneously: the amount and form of PI depend on the public competitive state surrounding the best bid and offer, so the same execution location can have different implications across quote-side venue states.

4.3 Price Improvement and Short-Horizon Mark-to-Midpoint Profits

We next ask whether trades that receive PI are associated with positive short-horizon trading outcomes for liquidity providers. We do not observe wholesalers' full inventories, hedges, or

[20] We use the ThetaData exchange identifiers attached to the displayed best bid and offer. PHLX includes the normalized PHLX/PHIL exchange code. As a data-quality check, the normalized PHLX-pair classification from ThetaData agrees with the corresponding SpiderRock exchange-side classification for 92.8% of trades.

netting across trades. The object we measure is therefore not firm-level profitability, but the mark-to-midpoint profit associated with each SLAN trade over short post-trade horizons.[21]

For each trade i at time t, we compute the trade-level profit as

$$\begin{aligned} Prof_{i,t}(\tau) &= -100 \cdot \text{Dir}_{i,t} \cdot \text{Volume}_{i,t} \cdot \left(P_{\text{midpoint},i,t+\tau} - P_{\text{actual},i,t}\right) \\ &= 100 \cdot \text{Dir}_{i,t} \cdot \text{Volume}_{i,t} \cdot \left(P_{\text{actual},i,t} - P_{\text{midpoint},i,t+\tau}\right), \end{aligned} \tag{4}$$

where $\text{Dir}_{i,t} = 1$ for buyer-initiated trades and $\text{Dir}_{i,t} = -1$ for seller-initiated trades, $\text{Volume}_{i,t}$ is the trade size, and τ indexes the horizon at which the position is evaluated. We consider horizons from the next available quote up to 120 seconds.[22]

Table 9 shows that the short-horizon mark-to-midpoint profit is positive across all horizons. When positions are evaluated at the next available quote, the average daily profit aggregated across SLAN trades is \$1.29 million. It remains economically large, at \$0.58 million, even at the 120-second horizon. Profitability is also broad-based: more than 50% of trades are profitable even at longer horizons, and the share of profitable volume exceeds 70% at the next available quote. These results indicate that, despite providing PI at execution, liquidity providers generate positive short-horizon mark-to-midpoint outcomes in aggregate. While the measure declines as the holding horizon lengthens, this pattern is broadly consistent with evidence from the stock market on the term structure of liquidity provision (Conrad and Wahal, 2020).

To understand the source of these profits, we decompose the return into an execution

[21] These liquidity-provider mark-to-midpoint outcomes differ from the customer-side gross-performance measure in Section 3.5.3. The latter abstracts from the opening execution-price component and evaluates signed changes in option value over longer horizons, whereas the present exercise includes the execution component and focuses on horizons up to 120 seconds.

[22] Our profit measure focuses on short-horizon price dynamics following each trade, capturing the immediate execution and adverse-selection component of profitability. In practice, wholesalers actively manage inventory risk and typically hedge their positions shortly after execution, especially with respect to delta exposure. This focus on short post-trade horizons is also consistent with Ni et al. (2021), who show that option-market-maker hedge rebalancing is an economically important channel through which option trading affects underlying-stock dynamics. As a result, even when positions are held over longer calendar horizons, the unhedged exposure associated with the initiating trade is often short-lived. The short horizons considered here therefore provide an approximation to the trade-level mark-to-midpoint outcome before inventory risk is fully neutralized.

component and a post-trade position component. Rewriting Equation (4), we obtain

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

The first term captures the execution gain or loss relative to the contemporaneous midpoint, while the second term captures the gain or loss from holding the acquired position over a short horizon.

This distinction is important for interpreting PI. A trade can involve a concession relative to the quoted bid or ask, which is what PI measures, while still generating a positive execution gain relative to the contemporaneous midpoint. Thus, the puzzle in Section 4.1 concerns why wholesalers are willing to offer substantial concessions relative to quoted terms, not whether they necessarily lose money relative to the midpoint at execution.

Table 10 reports this decomposition. Spread profit is large, averaging $1.326 million per day. This component is locked in at execution and reflects the pricing of the trade relative to the contemporaneous midpoint. By contrast, position profit is negative on average in the full sample and becomes increasingly negative as the holding horizon lengthens: it is -$36,082 at the next available quote, -$136,481 after 1 second, -$425,567 after 15 seconds, and -$745,577 after 120 seconds. Unlike spread profit, this component reflects post-trade midpoint movements rather than the execution price itself. The systematic decline over time indicates that post-trade prices tend to move against liquidity providers on average, consistent with adverse-selection effects being gradually realized at longer horizons. Thus, although aggregate short-horizon mark-to-midpoint profits remain positive because of the execution component, the positions acquired by liquidity providers are, on average, increasingly unfavorable when evaluated over longer post-trade windows.

The subsample results reveal a striking contrast: high-PI trades generate positive and increasing

position profits, while low-PI trades exhibit increasingly negative outcomes. For trades receiving at least 50% PI, position profit is positive at every horizon and increases over time, from $36,271 at the next quote to $141,118 after 1 second, $327,389 after 15 seconds, and $452,820 after 120 seconds. By contrast, for trades receiving less than 50% PI, position profit is negative throughout and becomes increasingly negative, ranging from -$72,353 at the next quote to -$277,599 after 1 second, -$752,956 after 15 seconds, and -$1.198 million after 120 seconds. As a result, the gap between the two groups widens sharply over time, reaching about $419 thousand at the 1-second horizon and about $1.65 million at 120 seconds.

This divergence reveals systematic differences in post-trade price dynamics across trades. Low-PI trades are associated with increasingly negative position outcomes, consistent with adverse-selection effects being gradually realized as prices move in the direction of the initiating order. High-PI trades, in contrast, are associated with positive and increasing position outcomes, suggesting that these trades are less exposed to adverse selection or reflect liquidity-driven order flow and temporary price pressure that subsequently reverses. Importantly, the high-PI group remains profitable even after these trades receive large execution concessions relative to the quoted bid or ask. In other words, for these trades, subsequent price movements more than offset the initial concession, and they do so almost immediately. In this sense, large PI is not simply a mechanical transfer to investors; it is associated with more favorable subsequent position outcomes.

This contrast helps to explain the puzzle in Section 4.1. If large PI merely reflects competitive pressure that compressed profits, then trades with larger concessions should perform worse after execution. Instead, the data show the opposite: trades receiving larger PI subsequently generate more favorable position outcomes, while trades receiving smaller concessions become increasingly costly to hold. This pattern suggests that wholesalers do not apply PI uniformly across trades, but instead use it selectively as part of a broader trade-off between execution pricing and position risk. Rather than relying solely on the quoted spread as a source of protection against adverse selection, wholesalers appear to balance immediate concessions against the expected profitability of the positions they acquire. In this sense, PI reflects a more active role for inventory management

and trade selection than in a purely spread-based view of market making.[23]

5 Robustness and Additional Model Specifications

Section 3.3 reports our main result using the one-month rolling GS-LASSO specification estimated within the five execution-price-location groups. We now examine whether this performance depends on a particular feature set, training horizon, or updating scheme. The goal is not to introduce another main model, but to verify that the classification gains are stable across reasonable implementation choices.

Table 11 reports out-of-sample accuracy across model specifications. Panel A estimates an unrestricted high-complexity XGBoost model using the full set of 445 candidate features, while Panel B reports GS-LASSO after SHAP-LASSO feature selection within the five execution-price-location groups. Within each panel, we compare three training horizons—one day, one week, and one month—with static and rolling updating schemes. A rolling specification re-estimates the model using the most recent training window before each evaluation period, whereas a static specification estimates the model once and applies the same fitted classifier throughout the entire out-of-sample period.

Several patterns emerge. First, feature selection does not sacrifice predictive performance relative to the unrestricted high-complexity benchmark. Under the one-month rolling specification, the all-feature high-complexity XGBoost model achieves 80.45% accuracy, while GS-LASSO with selected features achieves 80.46%. The same pattern holds across the other training horizons: the selected-feature models perform very similarly to the corresponding all-feature high-complexity models. This comparison is not intended to show that SHAP-LASSO screening mechanically improves accuracy relative to an already flexible unrestricted learner. Instead, it shows that the screening step preserves essentially all predictive information while substantially disciplining the

[23] Online Appendix I provides descriptive evidence that PI is related to model-implied directional signal strength and to the leading GS-LASSO signals.

predictor space. Thus, GS-LASSO delivers a more parsimonious and interpretable specification without sacrificing out-of-sample performance.[24]

Second, rolling specifications generally outperform static specifications with the same training horizon. For example, within the GS-LASSO specifications, the one-day rolling model achieves 79.02% accuracy compared with 77.55% for the one-day static model, the one-week rolling model achieves 80.05% compared with 78.50% for the one-week static model, and the one-month rolling model achieves 80.46% compared with 79.28% for the one-month static model. This pattern is consistent with the evidence in Section 3.2 that the price-improvement environment varies over time. Allowing the mapping from public market information to trade direction to update over time improves out-of-sample identification.

Third, the one-month rolling specification performs best among the GS-LASSO models. Table 12 reports pairwise comparisons of daily out-of-sample accuracy across GS-LASSO specifications. The one-month rolling model outperforms the one-week rolling model by 0.41 percentage points, with a t-statistic of 2.92, and it outperforms the one-month static, one-day rolling, one-week static, and one-day static specifications by 1.17, 1.44, 1.96, and 2.91 percentage points, respectively. These differences confirm that the one-month rolling design is the strongest implementation in our data. At the same time, the nearest alternative, the one-week rolling model, also achieves accuracy close to 80%, indicating that the main result is not an artifact of a narrowly chosen estimation window.

Taken together, these robustness results support the main interpretation of the paper. The gains from GS-LASSO are not driven by using all available predictors, by a single static training sample, or by a fragile choice of training horizon. Instead, the model performs well across a range of reasonable specifications, while the one-month rolling selected-feature implementation provides

[24] In our implementation, the unrestricted all-feature high-complexity XGBoost model takes about four times longer to train than the corresponding GS-LASSO specification. Online Appendix H reports all-feature low-complexity XGBoost benchmarks. The one-month rolling low-complexity benchmark achieves 79.23% accuracy, compared with 80.46% for GS-LASSO, indicating that the full GS-LASSO pipeline adds predictive value beyond the low-complexity first stage.

the best combination of feature discipline, model flexibility, and out-of-sample accuracy.

6 Conclusion

This paper studies the accuracy of the widely used midpoint rule for identifying option trade direction, one of the most fundamental inputs to empirical research on transaction costs, effective spreads, order imbalance, price impact, demand pressure, and information flow. Using a proprietary PHLX dataset with true initiator-side labels, we show that the widely used midpoint rule performs only slightly better than random classification in auction-based option trades. This finding identifies a first-order measurement failure: in modern option auctions, pervasive price improvement weakens the traditional mapping between execution price and trade direction.

We propose GS-LASSO, an interpretable machine-learning framework that represents a major step toward solving this failure of the midpoint rule. The framework combines XGBoost, SHAP-based interpretation, and LASSO feature selection to recover trade direction from the broader public information set. Rather than treating trade direction as a deterministic function of execution price, GS-LASSO views it as a state-dependent inference problem. Importantly, the framework does not discard the information contained in the midpoint. Instead, it learns when the execution-price signal is reliable and when it should be supplemented by information on market liquidity, depth, volatility, trade size, short-horizon price dynamics, and fair-value measures.

Empirically, GS-LASSO substantially improves trade classification, increasing the average out-of-sample accuracy from approximately 55% under the midpoint rule to 80.46% in our baseline specification. The improvement is not only statistically significant but also economically meaningful. It translates directly into more accurate empirical measures constructed from signed trades. In particular, GS-LASSO-implied order imbalance is substantially closer to the true order imbalance, while GS-LASSO-implied effective half-spreads greatly reduce the transaction-cost overstatement induced by midpoint-based trade classification. A same-sample gross-performance

application further shows that positive gross-performance estimates are sensitive to the direction measure used and become statistically insignificant under exchange-recorded initiator-side labels.

Accurate trade-direction measurement also changes our understanding of price improvement in auction-based option markets. Using exchange-recorded trade direction, we show that substantial price improvement can coexist with positive short-horizon mark-to-midpoint profits for liquidity providers and that execution concessions vary systematically with displayed quote conditions and subsequent position performance.

More broadly, our findings demonstrate that modern market design fundamentally challenges traditional empirical measurement methods. In auction-based option markets, execution prices alone are no longer sufficient to recover trade direction or accurately measure trading costs. Reliable inference instead requires conditioning on the broader public information environment surrounding each trade. We expect the framework developed here to be useful in future research on trading behavior across market venues, execution quality under alternative market designs, and the effects of market structure and regulation on retail trading outcomes.

References

Amaya, Diego, Pedro A. Garcia-Ares, Neil D. Pearson, and Aurelio Vasquez, 2025, New evidence on the performance of customer options trades.

Amihud, Yakov, and Haim Mendelson, 1980, Dealership market: Market-making with inventory, *Journal of Financial Economics* 8, 31–53.

Anand, Amber, and Dmitriy Muravyev, 2024, Does internalization impact quote competition?, Available at SSRN 4891227.

Avramov, Doron, Si Cheng, and Lior Metzker, 2023, Machine learning vs. economic restrictions: Evidence from stock return predictability, *Management Science* 69, 2587–2619.

Baldauf, Markus, Joshua Mollner, and Yueshen Bart Zhou, 2024, Siphoned apart: A portfolio perspective on order flow segmentation, *Journal of Financial Economics* 154, 103807.

Barardehi, Yashar H, Vincent Bogousslavsky, and Dmitriy Muravyev, 2025, Unpacking retail trading costs: the role of options trading and limit order usage, Available at SSRN 5979974.

Bartlett, Robert P., and Maureen O'Hara, 2024, Navigating the murky world of hidden liquidity, Available at SSRN 4988855.

Battalio, Robert H, Craig W Holden, Matthew Pierson, John J Shim, and Jun Wu, 2026, Latency and the look-ahead bias in trade and quote data, Available at SSRN 5907665.

Bogousslavsky, Vincent, Vyacheslav Fos, and Dmitriy Muravyev, 2024, Informed trading intensity, *The Journal of Finance* 79, 903–948.

Bryzgalova, Svetlana, Anna Pavlova, and Taisiya Sikorskaya, 2023, Retail trading in options and the rise of the big three wholesalers, *The Journal of Finance* 78, 3465–3514.

Bühlmann, Peter, and Torsten Hothorn, 2007, Boosting algorithms: Regularization, prediction and model fitting, *Statistical Science* 22, 477–505.

Chen, Luyang, Markus Pelger, and Jason Zhu, 2024, Deep learning in asset pricing, *Management Science* 70, 714–750.

Chen, Tianqi, and Carlos Guestrin, 2016, XGBoost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chinco, Alex, Adam D. Clark-Joseph, and Mao Ye, 2019, Sparse signals in the cross-section of returns, *The Journal of Finance* 74, 449–492.

Conrad, Jennifer, and Sunil Wahal, 2020, The term structure of liquidity provision, *Journal of Financial Economics* 136, 239–259.

Coval, Joshua D., and Tyler Shumway, 2001, Expected option returns, *The Journal of Finance* 56, 983–1009.

Easley, David, Marcos López de Prado, Maureen O'Hara, and Zhibai Zhang, 2021, Microstructure in the machine age, *The Review of Financial Studies* 34, 3316–3363.

Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.

Eaton, Gregory W, T Clifton Green, Brian S Roseman, and Yanbin Wu, 2026, Retail option traders and the implied volatility surface, *Journal of Financial Economics* 177, 104238.

Ellis, Katrina, Roni Michaely, and Maureen O'Hara, 2000, The accuracy of trade classification rules: Evidence from Nasdaq, *Journal of Financial and Quantitative Analysis* 35, 529–551.

Ernst, Thomas, and Chester Spatt, 2026, Payment for order flow and option internalization, *The Review of Financial Studies* Published online February 20, 2026.

Ernst, Thomas, Chester Spatt, and Jian Sun, 2025, Would order-by-order auctions be competitive?, *The Journal of Finance* 80, 1879–1927.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the factor zoo: A test of new factors, *The Journal of Finance* 75, 1327–1370.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *The Review of Financial Studies* 33, 2326–2377.

Friedman, Jerome H., 2001, Greedy function approximation: A gradient boosting machine, *The Annals of Statistics* 29, 1189–1232.

Garman, Mark B., 1976, Market microstructure, *Journal of Financial Economics* 3, 257–275.

Giglio, Stefano, Yuan Liao, and Dacheng Xiu, 2021, Thousands of alpha tests, *The Review of Financial Studies* 34, 3456–3496.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Grauer, Caroline, Philipp Schuster, and Marliese Uhrig-Homburg, 2026, Unmasking option demand: New classification methods and stock return predictability, Working paper.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *The Review of Financial Studies* 33, 2223–2273.

Han, Yufeng, Ai He, David E. Rapach, and Guofu Zhou, 2024, Cross-sectional expected returns: New Fama–MacBeth regressions in the era of machine learning, *Review of Finance* 28, 1807–1831.

Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *The Journal of Finance* 46, 179–207.

Hendershott, Terrence, Saad Ali Khan, and Ryan Riordan, 2026, Option auctions, *The Review of Financial Studies* 39, 783–834.

Ho, Thomas, and Hans R. Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics* 9, 47–73.

Hu, Jianfeng, 2014, Does option trading convey stock price information?, *Journal of Financial Economics* 111, 625–645.

Huang, Roger D., and Hans R. Stoll, 1997, The components of the bid-ask spread: A general approach, *The Review of Financial Studies* 10, 995–1034.

Kelly, Bryan, Semyon Malamud, and Kangying Zhou, 2024, The virtue of complexity in return prediction, *The Journal of Finance* 79, 459–503.

Kelly, Bryan T., and Semyon Malamud, 2025, Understanding the virtue of complexity, Swiss Finance Institute Research Paper 25-96, Swiss Finance Institute.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

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

Light, Nathaniel, Denys Maslov, and Oleg Rytchkov, 2017, Aggregation of information about the cross section of stock returns: A latent variable approach, *The Review of Financial Studies* 30, 1339–1381.

Lundberg, Scott M., and Su-In Lee, 2017, A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems* 30.

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

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

Newey, Whitney K, and Kenneth D West, 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, 631–653.

Ni, Sophie X, Neil D Pearson, Allen M Poteshman, and Joshua White, 2021, Does option trading have a pervasive impact on underlying stock prices?, *The Review of Financial Studies* 34, 1952–1986.

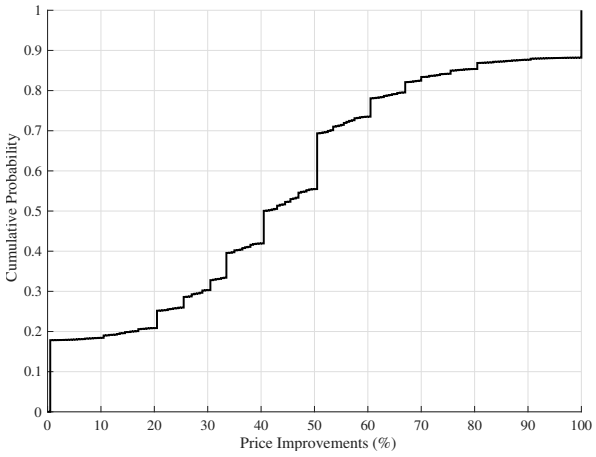
Savickas, Robert, and Arthur J. Wilson, 2003, On inferring the direction of option trades, *Journal of Financial and Quantitative Analysis* 38, 881–902.

Zhan, Xintong, Bing Han, Jie Cao, and Qing Tong, 2022, Option return predictability, *The Review of Financial Studies* 35, 1394–1442.

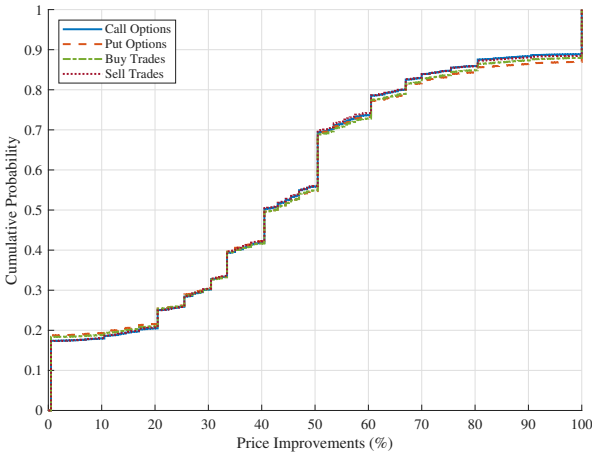
Zhang, Tong, and Bin Yu, 2005, Boosting with early stopping: Convergence and consistency, *The Annals of Statistics* 33, 1538–1579.

Figure 1. Empirical Cumulative Distribution Functions of Price Improvement

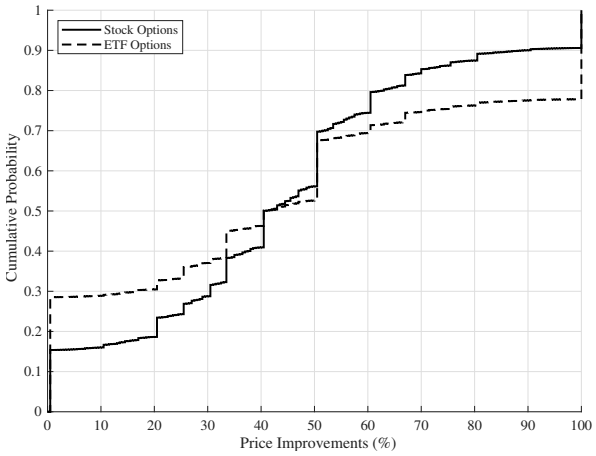
This figure plots empirical cumulative distribution functions of price improvement, defined in Equation (1). Panel A reports the full sample. Panel B separates call and put options as well as buy and sell trades. Panel C separates stock and ETF options. Panel D separates trades by whether the displayed national best bid and/or national best offer are posted on PHLX. Price improvement is measured as a percentage of the quoted spread. The sample includes all single-leg auction, non-ISO (SLAN) option trades written on stocks and ETFs from January 1, 2024 to December 31, 2024.



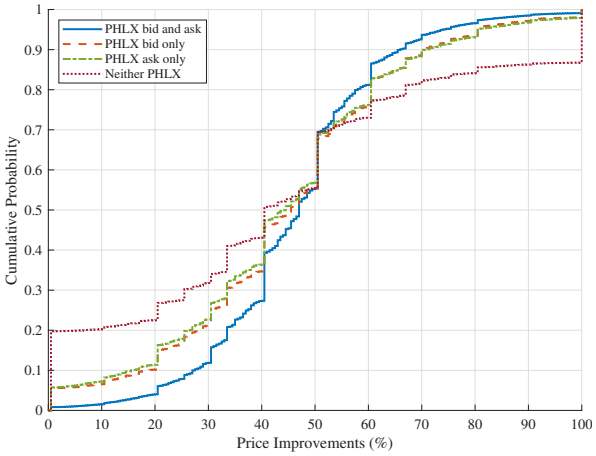
Panel A. Full sample



Panel B. Call/put and buy/sell trades



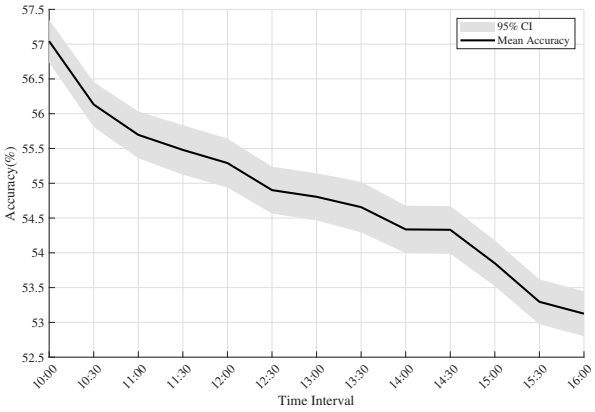
Panel C. Stock and ETF options



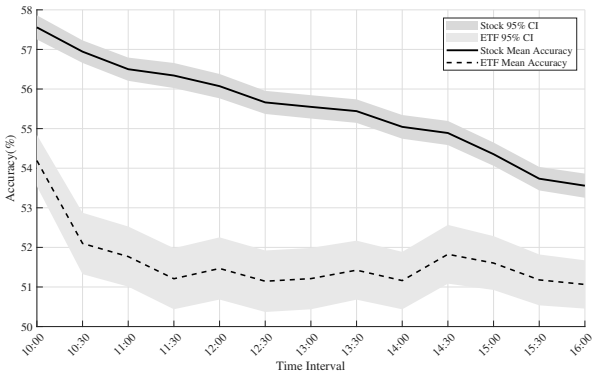
Panel D. PHLX quote-side venue

Figure 2. Intraday Accuracy of the Midpoint Rule

This figure reports intraday midpoint-rule accuracy. Panel A reports accuracy for the full sample by half-hour interval. Panel B reports accuracy separately for stock and ETF options. Shaded areas denote 95% confidence intervals across trading days. The sample includes all single-leg auction, non-ISO (SLAN) option trades written on stocks and ETFs from January 1, 2024 to December 31, 2024.



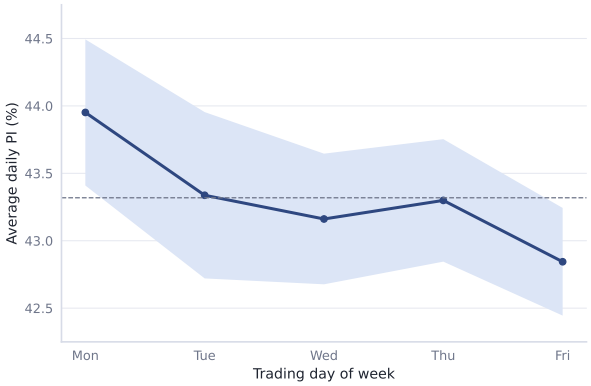
Panel A. Overall



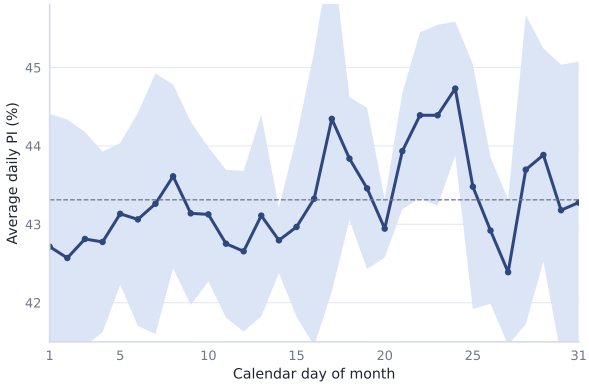
Panel B. Stock and ETF options

Figure 3. Calendar Patterns in Price Improvement

This figure reports calendar patterns in price improvement. Panel A reports average daily price improvement by trading day of week, with 95% confidence intervals across days. Panel B aligns observations by calendar day of month and reports average daily price improvement, with 95% confidence intervals across observed months. The sample includes SLAN option trades written on stocks and ETFs in 2024.



Panel A. Weekday pattern



Panel B. Calendar-day pattern

Figure 4. Time-Series Accuracy of GS-LASSO and the Midpoint Rule

This figure plots daily out-of-sample classification accuracy for the GS-LASSO specification and the midpoint rule on the same prediction panel. The GS-LASSO model is trained using trades from month m and evaluated on trades from month $m+1$. The out-of-sample period runs from February 1, 2024 to December 31, 2024 and includes SLAN option trades written on stocks and ETFs.

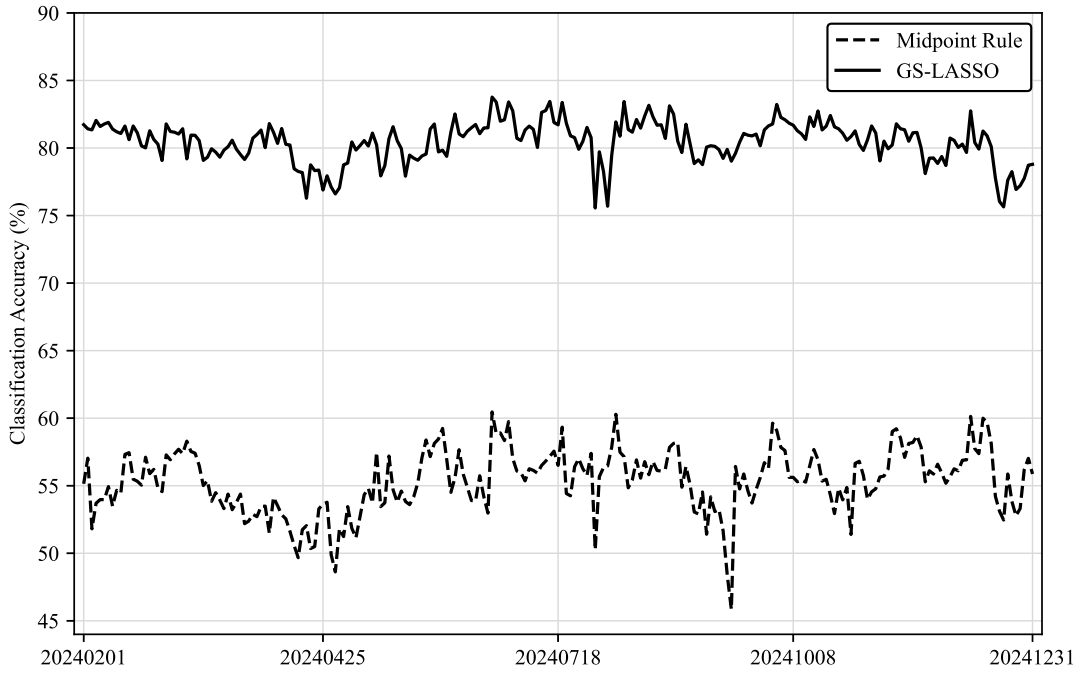


Figure 5. Annual Average SHAP Importance

This figure reports the ten largest annual mean absolute SHAP values for the GS-LASSO specification. For each feature, SHAP importance is computed within each price-location model and monthly training window, weighted within the window by the corresponding training-sample share, and averaged across the 11 rolling windows. The figure reports SHAP importance from the first-stage models used for feature attribution and screening. Feature definitions are provided in Online Appendix M.

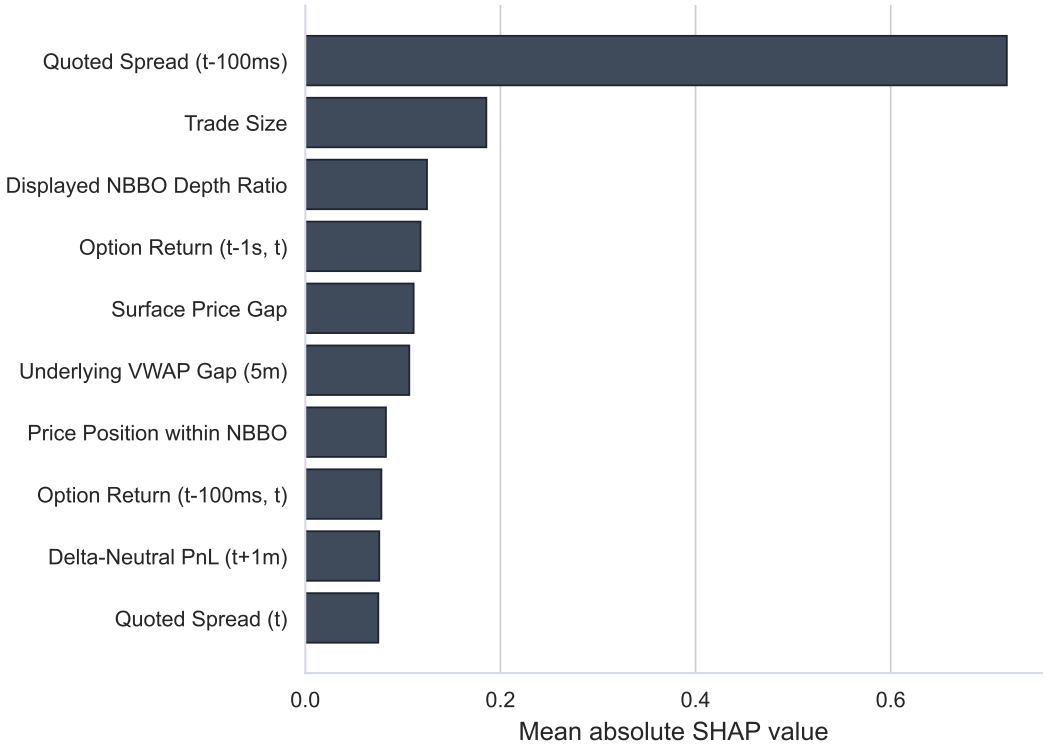
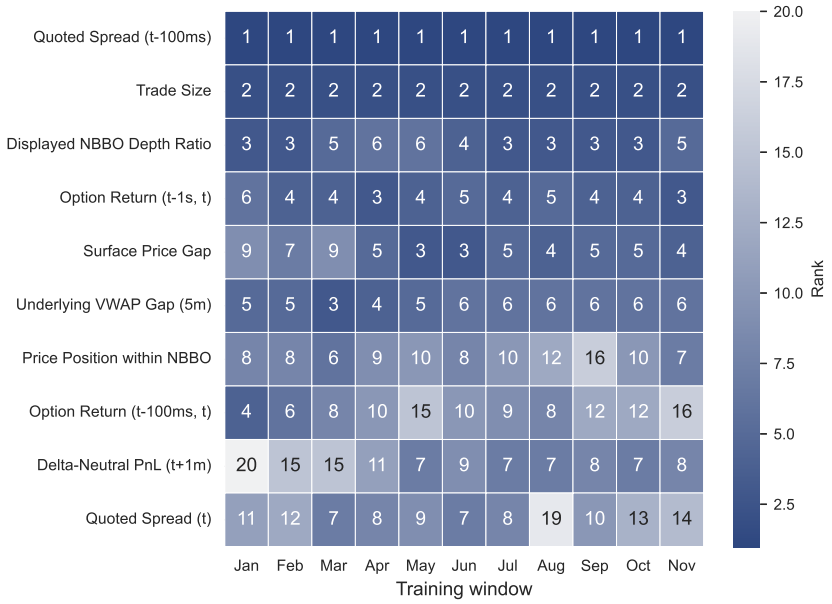
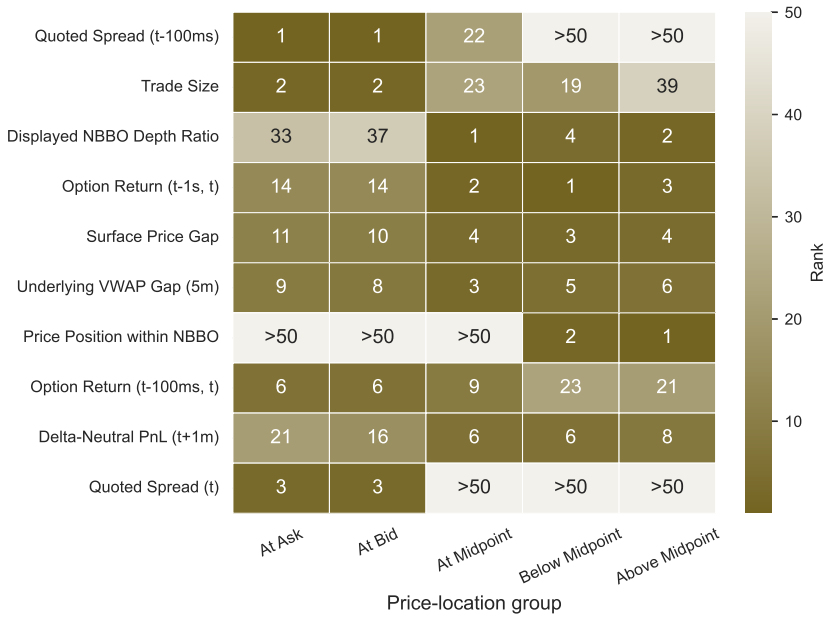


Figure 6. SHAP Feature-Importance Ranks Across Time and Price-Location Groups

This figure reports rank heatmaps for the ten features shown in Figure 5. Panel A ranks these features within each monthly rolling training window. Panel B ranks them within each of the five execution-price-location groups: at ask, at bid, at midpoint, below midpoint, and above midpoint. Smaller numbers indicate greater SHAP importance; cells marked > 50 indicate that the feature is outside the top 50 for that window or group.



Panel A. Rolling monthly windows



Panel B. Five execution-price-location groups

Table 1 Summary Statistics of Daily SLAN Trade Characteristics

This table reports summary statistics of daily single-leg auction, non-ISO (SLAN) option trade characteristics. The sample includes all SLAN option trades written on stocks and ETFs from January 1, 2024 to December 31, 2024. Trade count is reported in thousands and volume in millions. We also report trading volumes by moneyness (ATM, ITM, and OTM) and by underlying asset class (stock and ETF options). Trade size is the daily average across all trades. Days to expiration is calculated as the trade-size-weighted average time to expiration (in years), multiplied by 365. The percentage quoted spread is the daily trade-size-weighted average quoted spread. The call trade percentage and buy trade percentage measure the proportions of call option trades and buyer-initiated trades, respectively.

	Mean	Std	Min	5%	25%	Median	75%	95%	Max
Trade Count (Thousand)	185.24	37.21	57.91	131.81	162.71	181.05	204.98	258.23	317.34
Volume (Million)	1.57	0.30	0.50	1.12	1.37	1.55	1.76	2.10	2.38
ATM Volume (Million)	0.96	0.21	0.29	0.65	0.82	0.95	1.11	1.30	1.62
ITM Volume (Million)	0.22	0.05	0.08	0.15	0.19	0.21	0.25	0.31	0.42
OTM Volume (Million)	0.38	0.09	0.13	0.26	0.32	0.37	0.43	0.54	0.73
Stock Option Volume (Million)	1.14	0.24	0.37	0.79	0.96	1.11	1.28	1.59	1.80
ETF Option Volume (Million)	0.43	0.09	0.13	0.29	0.37	0.43	0.49	0.59	0.68
Trade Size	8.49	0.64	6.82	7.41	8.09	8.46	8.87	9.56	10.37
Days to Expiration	31.54	3.91	22.73	25.71	28.94	31.34	33.87	38.30	48.74
Percentage Quoted Spread	0.1509	0.0260	0.1022	0.1145	0.1314	0.1481	0.1677	0.1963	0.2470
Call Trade Percentage	65.00%	2.95%	58.33%	60.20%	62.78%	64.95%	67.33%	69.75%	71.86%
Buy Trade Percentage	50.61%	1.71%	44.72%	47.69%	49.87%	50.83%	51.71%	53.00%	54.06%

Table 2 Accuracy of Public Trade-Signing Rules

This table reports summary statistics of daily classification accuracy for two public trade-signing rules, evaluated relative to the exchange-recorded true direction. For each trading day, accuracy is measured as the percentage of correctly classified trades. Panel A reports the midpoint rule. Panel B reports the exchange-quote refinement of Muravyev (2016), which applies the midpoint rule away from the NBBO midpoint and uses the local exchange midpoint for exact-NBBO-midpoint trades. Results are reported for all trades and separately for buy trades, sell trades, call options, and put options. The sample includes SLAN option trades written on stocks and ETFs from January 1, 2024 to December 31, 2024.

	Obs	Mean	Std	Min	5%	25%	Median	75%	95%	Max
Panel A. Midpoint rule										
All Trades	252	55.34%	2.39%	45.85%	51.19%	53.85%	55.59%	57.03%	59.05%	60.47%
Buy Trades	252	54.75%	2.66%	44.34%	50.00%	53.24%	55.07%	56.49%	58.86%	60.18%
Sell Trades	252	55.95%	2.27%	47.32%	52.32%	54.56%	56.11%	57.47%	59.42%	61.25%
Call Trades	252	55.67%	2.45%	46.89%	50.95%	54.20%	55.92%	57.34%	59.17%	60.85%
Put Trades	252	54.60%	2.71%	43.92%	50.12%	52.96%	54.49%	56.66%	58.67%	62.64%
Panel B. Muravyev (2016) exchange-quote refinement										
All Trades	252	60.16%	2.36%	50.80%	55.83%	58.73%	60.24%	61.89%	63.70%	65.07%
Buy Trades	252	59.46%	2.62%	49.33%	54.57%	57.91%	59.76%	61.16%	63.35%	65.18%
Sell Trades	252	60.86%	2.26%	52.24%	57.27%	59.34%	61.03%	62.40%	64.28%	65.81%
Call Trades	252	60.37%	2.42%	51.69%	55.90%	58.99%	60.58%	62.07%	63.89%	65.52%
Put Trades	252	59.62%	2.69%	49.14%	55.24%	57.95%	59.53%	61.61%	64.02%	66.71%

Table 3 Summary Statistics of SLAN Price Improvements

This table reports summary statistics of price improvement, defined in Equation (1). Panel A reports the full sample and splits by trade direction and option type. Panel B reports the five execution-price-location groups used in the GS-LASSO implementation: trades at the ask, at the bid, at the midpoint, inside the spread below the midpoint, and inside the spread above the midpoint. Panel C reports results by whether PHLX posts the displayed national best bid and/or national best offer. The sample includes SLAN option trades written on stocks and ETFs from January 1, 2024 to December 31, 2024.

	Obs	Mean	Std	5%	25%	Median	75%	95%
Panel A. Full sample and basic trade characteristics								
All Trades	46.68M	43.23%	30.27%	0.00%	20.00%	40.00%	60.00%	100.00%
Buy Trades	23.22M	43.46%	30.58%	0.00%	20.00%	42.22%	60.00%	100.00%
Sell Trades	23.47M	43.00%	29.96%	0.00%	20.83%	40.00%	60.00%	100.00%
Call Trades	30.54M	43.07%	29.86%	0.00%	20.00%	40.00%	60.00%	100.00%
Put Trades	16.14M	43.53%	31.04%	0.00%	20.00%	42.11%	60.00%	100.00%
Panel B. Five execution-price-location groups								
At Ask	6.98M	38.86%	48.74%	0.00%	0.00%	0.00%	100.00%	100.00%
At Bid	6.86M	40.77%	49.14%	0.00%	0.00%	0.00%	100.00%	100.00%
At Midpoint	6.47M	50.00%	0.00%	50.00%	50.00%	50.00%	50.00%	50.00%
Below Midpoint	13.52M	43.19%	18.90%	16.67%	30.00%	40.00%	60.00%	80.00%
Above Midpoint	12.85M	43.54%	18.37%	20.00%	30.00%	40.00%	60.00%	77.78%
Panel C. Displayed PHLX quote-side state								
PHLX Bid and Ask	1.01M	46.14%	16.15%	20.00%	37.14%	46.67%	54.07%	73.85%
PHLX Bid Only	2.48M	44.30%	21.41%	0.00%	30.00%	45.00%	57.69%	80.00%
PHLX Ask Only	2.48M	43.74%	21.98%	0.00%	30.00%	44.00%	57.14%	80.00%
Neither Side on PHLX	40.72M	43.06%	31.41%	0.00%	20.00%	40.00%	60.00%	100.00%

Table 4 Summary Statistics of GS-LASSO Accuracy

This table reports summary statistics of daily out-of-sample classification accuracy for the GS-LASSO specification. For each trading day, accuracy is measured as the percentage of correctly classified trades relative to the exchange-recorded true direction. Results are reported for all trades and separately for buy trades, sell trades, call options, and put options. The out-of-sample period runs from February 1, 2024 to December 31, 2024 and includes SLAN option trades written on stocks and ETFs.

	Obs	Mean	Std	Min	5%	25%	Median	75%	95%	Max
All Trades	231	80.46%	1.54%	75.56%	77.68%	79.64%	80.63%	81.48%	82.75%	83.77%
Buy Trades	231	80.14%	1.67%	74.15%	77.28%	79.31%	80.12%	81.17%	82.75%	85.13%
Sell Trades	231	80.75%	1.90%	74.78%	77.17%	79.62%	80.99%	82.04%	83.36%	84.71%
Call Trades	231	80.20%	1.52%	75.06%	77.45%	79.41%	80.42%	81.15%	82.54%	83.60%
Put Trades	231	80.94%	1.70%	75.12%	77.59%	80.05%	81.12%	82.09%	83.43%	84.73%

Table 5 GS-LASSO Performance Across Market Conditions

This table reports out-of-sample classification accuracy for the GS-LASSO specification and the midpoint rule across subsamples defined by contract and trading characteristics. The out-of-sample test includes all single-leg auction, non-ISO (SLAN) option trades written on stocks and ETFs from February 1, 2024 to December 31, 2024. For each characteristic, the sample is split at its median into below-median and above-median groups. The "Difference" column reports GS-LASSO accuracy minus midpoint-rule accuracy. Standard errors, expressed in percentage points, are time-series standard errors of the paired daily accuracy differences across the 231 out-of-sample trading days. Significance is based on two-sided one-sample t-tests of whether the mean daily difference equals zero. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels. Midpoint Effective Spread is $|P_i - M_i|/M_i$, where P_i is the execution price and M_i is the contemporaneous NBBO midpoint.

	Below Median			Above Median		
Characteristics	**GS-LASSO**	**Midpoint**	**Difference**	**GS-LASSO**	**Midpoint**	**Difference**
Trade Size	80.25%	56.39%	23.86%***	80.77%	53.80%	26.97%***
			(0.13)			(0.16)
Option Price	84.25%	51.87%	32.38%***	76.62%	58.81%	17.81%***
			(0.23)			(0.10)
Quoted Spread (Dollar)	85.00%	54.74%	30.26%***	75.52%	56.09%	19.43%***
			(0.22)			(0.15)
Midpoint Effective Spread	75.87%	47.25%	28.62%***	85.05%	63.31%	21.74%***
			(0.11)			(0.21)
Trade–Quote Time Diff	82.25%	56.81%	25.44%***	78.61%	53.81%	24.80%***
			(0.16)			(0.12)
Time to Expiration	82.28%	54.91%	27.37%***	78.61%	55.77%	22.84%***
			(0.17)			(0.11)
Moneyness	80.45%	55.18%	25.27%***	80.47%	55.49%	24.98%***
			(0.15)			(0.14)
Bid-Ask Depth Ratio	80.51%	55.28%	25.23%***	80.40%	55.38%	25.02%***
			(0.14)			(0.15)

Table 6 Accuracy of Count-Based Order Imbalance Measurement

This table compares count-based order imbalance constructed using GS-LASSO-implied direction and midpoint-implied direction against order imbalance constructed using the exchange-recorded true direction. For each underlying ticker and trading day, count-based order imbalance is the average signed direction across SLAN trades written on that underlying. Underlying-ticker groups are ranked by total out-of-sample option volume. Volume Share reports each group's share of total out-of-sample option volume. Panel A reports the average daily cross-sectional correlation with true order imbalance, where Improvement is defined as the GS-LASSO correlation minus the midpoint correlation. Panel B reports the mean squared error relative to true order imbalance, where Improvement is defined as one minus the GS-LASSO mean squared error divided by the midpoint mean squared error.

Underlying Ticker Group	Volume Share	GS-LASSO	Midpoint	Improvement
Panel A. Correlation with true order imbalance				
Top 20 Underlying Tickers	56.46%	0.8831	-0.2140	1.0970
Top 100 Underlying Tickers	77.45%	0.8379	0.2253	0.6125
Top 500 Underlying Tickers	94.49%	0.7597	0.3660	0.3937
Overall	100.00%	0.6103	0.2806	0.3297
Panel B. Mean squared error of order imbalance				
Top 20 Underlying Tickers	56.46%	0.0028	0.0229	88.0%
Top 100 Underlying Tickers	77.45%	0.0084	0.0341	75.2%
Top 500 Underlying Tickers	94.49%	0.0386	0.0948	59.3%
Overall	100.00%	0.3324	0.6050	45.1%

Table 7 Effective Half-Spreads under Alternative Direction Inputs

This table reports effective half-spreads defined in Equation (3) for SLAN trades in the February–December 2024 out-of-sample period. Panel A reports $100 \times ES_i^k / M_i$, expressed as a percentage of the quote midpoint. Panel B reports $100 \times ES_i^k$, expressed in cents of option premium. For each day, statistics are computed from trade-level observations and then summarized across all trading days.

Direction Input	Mean	Std.	Median
Panel A. Effective spread as percent of midpoint			
True	0.75%	0.27%	0.72%
GS-LASSO	1.09%	0.27%	1.04%
Midpoint	3.06%	0.51%	2.99%
Panel B. Effective spread in cents			
True	0.83	0.25	0.77
GS-LASSO	1.32	0.35	1.27
Midpoint	3.11	0.54	3.04

Table 8 Trade Direction and Gross-Performance Inference

This table reports mean daily gross performance, in millions of dollars, for PHLX SLAN trades from February 1 through December 31, 2024. Positive values denote gross performance to the initiating SLAN side. Muravyev (2016) Quote-Based Direction implements the two-stage quote procedure of Muravyev (2016) used by Bryzgalova et al. (2023) in their gross-performance analysis, whereas Exchange-Recorded Direction uses the initiator side recorded by PHLX. Within each holding horizon, the same trades, capped trade sizes, initial and terminal valuation prices, sample filters, and aggregation procedure are used across columns; only the direction input changes. Newey–West standard errors are reported in parentheses below the estimates. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Holding Horizon	Muravyev (2016) Quote-Based Direction	Exchange-Recorded Direction
Intraday	2.385***	−0.118
	(0.178)	(0.171)
1 Day	2.683***	−0.276
	(0.304)	(0.337)
2 Days	2.506***	0.090
	(0.419)	(0.629)
5 Days	2.878***	−0.007
	(0.655)	(0.651)
10 Days	3.059***	0.236
	(1.006)	(0.961)
Expiration	1.397	−0.215
	(0.985)	(1.193)

Table 9 Short-Horizon Mark-to-Midpoint Profits Across Holding Horizons

This table reports short-horizon mark-to-midpoint profits for SLAN trades under alternative position-evaluation horizons. Profits are computed using the transaction price, exchange-recorded true direction, trade size, and the option midpoint at the evaluation horizon. Average Daily Profit denotes the total profit across all trades within a day, averaged across trading days. Profit per trade is reported in dollars per contract. Profit Trade % denotes the percentage of trades with positive profits, and Profit Volume % denotes the percentage of trading volume associated with profitable trades. The sample includes SLAN option trades written on stocks and ETFs from January 1, 2024 to December 31, 2024.

Holding Horizon	Average Daily Profit	Profit Per Trade	Profit Trade %	Profit Volume %
Next Quote	$1,290,105	$6.96	56.51%	70.81%
1 Second	$1,189,706	$6.42	55.37%	68.28%
5 Seconds	$1,063,515	$5.74	53.74%	63.20%
15 Seconds	$900,621	$4.86	52.58%	58.34%
30 Seconds	$768,084	$4.15	51.92%	55.64%
60 Seconds	$672,472	$3.63	51.38%	53.77%
90 Seconds	$662,820	$3.58	51.13%	53.05%
120 Seconds	$580,611	$3.13	50.98%	52.52%

Table 10 Decomposition of Short-Horizon Mark-to-Midpoint Profits

This table decomposes short-horizon mark-to-midpoint profits into spread and position components, as defined in Equation (5). Spread Profit is the execution gain relative to the contemporaneous midpoint. Position Profit is the gain or loss from holding the position until the evaluation horizon. Position Profit is reported for the full sample and separately for trades with price improvement greater than or equal to 50% and less than 50%. Average Daily Profit denotes the total profit across all trades within a day, averaged across trading days. The sample includes SLAN option trades written on stocks and ETFs from January 1, 2024 to December 31, 2024.

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

Table 11 Classification Accuracy Across Model Specifications

This table reports summary statistics of daily out-of-sample classification accuracy across model specifications. Panel A reports unrestricted high-complexity XGBoost models estimated with the full set of 445 candidate features. Panel B reports GS-LASSO specifications after SHAP-LASSO feature selection within the five execution-price-location groups. Within each panel, we compare six configurations formed by crossing three training horizons (1 day, 1 week, and 1 month) with two updating schemes (static and rolling). The available evaluation dates begin on January 3, 2024 for one-day models, January 9, 2024 for one-week models, and February 1, 2024 for one-month models; all evaluation periods end on December 31, 2024 and include SLAN option trades written on stocks and ETFs.

	Mean	Std	Min	5%	25%	Median	75%	95%	Max
Panel A. All features: high-complexity XGBoost									
1 Day Static	77.54%	2.01%	71.25%	74.19%	76.33%	77.72%	78.75%	80.90%	83.61%
1 Day Rolling	79.04%	1.67%	73.53%	76.27%	78.05%	79.15%	80.11%	81.70%	83.72%
1 Week Static	78.42%	1.86%	72.63%	75.17%	77.29%	78.55%	79.57%	81.16%	84.53%
1 Week Rolling	80.06%	1.54%	74.72%	77.39%	79.25%	80.31%	81.02%	82.43%	84.39%
1 Month Static	79.26%	1.64%	74.05%	76.33%	78.26%	79.49%	80.32%	81.66%	82.86%
1 Month Rolling	80.45%	1.54%	75.50%	77.74%	79.61%	80.68%	81.48%	82.71%	83.86%
Panel B. GS-LASSO selected features									
1 Day Static	77.55%	2.00%	71.34%	74.11%	76.31%	77.68%	78.75%	80.77%	83.58%
1 Day Rolling	79.02%	1.67%	73.47%	76.27%	78.02%	79.16%	80.07%	81.59%	83.73%
1 Week Static	78.50%	1.87%	72.63%	75.16%	77.41%	78.68%	79.66%	81.23%	84.47%
1 Week Rolling	80.05%	1.54%	74.76%	77.43%	79.17%	80.24%	81.04%	82.47%	84.35%
1 Month Static	79.28%	1.64%	74.00%	76.33%	78.28%	79.51%	80.38%	81.68%	82.87%
1 Month Rolling	80.46%	1.54%	75.56%	77.68%	79.64%	80.63%	81.48%	82.75%	83.77%

Table 12 Out-of-Sample Performance Across GS-LASSO Model Specifications

This table reports pairwise comparisons of out-of-sample daily classification accuracy across GS-LASSO model specifications. The models include one-month rolling, one-week rolling, one-month static, one-day rolling, one-week static, and one-day static specifications. Models are ordered by their average daily out-of-sample accuracy, so each reported upper-triangular entry is positive. Each cell reports the difference in average daily accuracy, defined as the row model minus the column model. The corresponding t-statistic is reported in parentheses. Asterisks represent statistical significance at the 1% (***), 5% (**), and 10% (*) levels. The available evaluation dates begin on January 3, 2024 for one-day models, January 9, 2024 for one-week models, and February 1, 2024 for one-month models; all evaluation periods end on December 31, 2024 and include SLAN option trades written on stocks and ETFs. Because the available dates differ across training horizons, the reported t-statistics are based on two-sided Welch unequal-variance two-sample tests using all available daily accuracy observations for each specification; observations are not matched by trading date.

	1W Rolling	1M Static	1D Rolling	1W Static	1D Static
1M Rolling	0.41%***	1.17%***	1.44%***	1.96%***	2.91%***
	(2.92)	(7.91)	(9.82)	(12.54)	(17.95)
1W Rolling		0.76%***	1.02%***	1.55%***	2.50%***
		(5.21)	(7.11)	(10.04)	(15.60)
1M Static			0.26%*	0.79%***	1.74%***
			(1.74)	(4.90)	(10.45)
1D Rolling				0.52%***	1.48%***
				(3.30)	(8.97)
1W Static					0.95%***
					(5.48)

Online Appendix

Inferring Option Trade Direction with Machine Learning

This online appendix accompanies the paper *Inferring Option Trade Direction with Machine Learning*.

A Institutional Details of PHLX Single-Leg Auction Mechanisms

The institutional setting for listed equity options differs from that of NMS stocks in a way that is central for understanding auction-based execution. Standardized listed options are exchange-traded products, and trading them off an exchange has historically been difficult in practice due to clearing and related legal and operational constraints.[25] As a result, when brokers route retail option orders to broker-dealer wholesalers (often in connection with payment for order flow arrangements), the wholesaler typically competes for execution by submitting the order into an exchange auction mechanism rather than executing the trade off-exchange in the manner commonly observed in equities.[26]

PHLX single-leg auctions implement this competition through short exposure periods and sealed responses. For the auction mechanisms in Nasdaq PHLX Options 3, Section 11, the exchange disseminates a broadcast message to members and permits members to submit electronic responses.[27] These responses represent non-firm interest: they can be canceled or modified at any

[25] See, e.g., SEC, "Special Study: Payment for Order Flow and Internalization in the Securities Markets" (1997), noting that while no rule explicitly requires listed options to trade only on an exchange, a combination of OCC bylaws and state laws makes off-exchange trading difficult.

[26] See, e.g., SEC, Regulation Best Execution (proposing release), which discusses wholesalers in the listed options market and the prevalence of exchange-run price improvement auctions and related internalization on options exchanges.

[27] Nasdaq PHLX Rules, Options 3, §11 (definition of "broadcast message" and "Response").

time prior to execution and are not displayed to any market participants.[28] The response window is exchange-designated but tightly bounded, with the time given for responses no less than 100 milliseconds and no more than one second.[29] This design makes participation highly competitive while limiting information leakage: during the exposure period, each participant knows its own submitted interest but does not observe the full set of competing responses.

Clearing occurs at the end of the exposure period, which further disciplines the competitive process. In the Block Order Mechanism, after the response window ends, the exchange either executes the order automatically or cancels it.[30] In the Facilitation Mechanism, upon entry of an agency order the exchange broadcasts the series, price, and size and accepts responses; at the end of the response period the facilitation order is automatically executed.[31] The Solicited Order Mechanism similarly broadcasts the proposed execution price and accepts responses, and then the agency order is automatically executed in full or cancelled at the end of the response period.[32] Trading halts constitute a salient exception: if a halt is initiated after an order is entered, these auctions are terminated without execution.[33] Otherwise, the mechanisms commit participants to compete over a fixed exposure interval rather than allowing an endogenous, discretionary early stop.

The price improvement auction on PHLX, PIXL, follows the same exposure logic but is governed separately under Options 3, Section 13 and is explicit about informational non-display. To initiate a PIXL auction, the initiating member submits a two-sided package consisting of an agency-side "PIXL Order" and a contra-side "Initiating Order," and specifies a "stop price" (and optionally an auto-match instruction).[34] The exchange then disseminates a PIXL Auction Notification (PAN) over the PHLX data feed.[35] Any person or entity may submit responses to the

[28] Nasdaq PHLX Rules, Options 3, §11 (responses are non-firm, cancellable/modifiable prior to execution, and not displayed).
[29] Nasdaq PHLX Rules, Options 3, §11 (response window bounds).
[30] Nasdaq PHLX Rules, Options 3, §11(a)(2).
[31] Nasdaq PHLX Rules, Options 3, §11(b)(2) and §11(b)(4).
[32] Nasdaq PHLX Rules, Options 3, §11(d)(2)–(3).
[33] Nasdaq PHLX Rules, Options 3, §11(a)(3), §11(b)(5), and §11(d)(4).
[34] Nasdaq PHLX Rules, Options 3, §13(b)(1)(A) (stop price and related instructions).
[35] Nasdaq PHLX Rules, Options 3, §13(b)(1)(C).

PAN.[36] Crucially, PAN responses are not visible to auction participants and are not disseminated to the Options Price Reporting Authority (OPRA).[37] PAN responses may be modified or cancelled during the auction, and responses inferior to the stop price are cancelled at the conclusion of the auction.[38] In equilibrium, these provisions imply that the competitive state of the auction is largely unobservable while the auction is ongoing, and market participants learn the outcome through the subsequent trade report rather than through real-time display of rival quotes.[39]

Although PHLX auctions are designed to run for a short, predetermined exposure period, PIXL specifies limited early-conclusion triggers. The PIXL auction concludes at the end of the auction period but may also conclude earlier under rule-defined events, including when the internal PBBO crosses the stop price on the same side of the market (for non-complex PIXL) or when a trading halt occurs.[40] At the same time, unrelated market or marketable interest arriving during the auction does not cause the PIXL auction to end early; such interest executes outside the auction and only participates in allocation if it remains at the time the auction ends.[41] This structure sustains a highly competitive environment while limiting prolonged exposure of the order and the response book.

The "non-ISO" nature of our single-leg auction notifications is important for interpretation because it governs whether execution must respect protected quotations. PHLX rules explicitly define ISO variants for certain auction mechanisms. For example, a "Facilitation ISO" (and a "PIXL ISO") is defined as a crossing submission "without regard for better priced Protected Bids/Offers" because the submitting member simultaneously routes ISOs to sweep superior protected interest.[42] By contrast, the baseline (non-ISO) auction mechanisms impose explicit

[36] Nasdaq PHLX Rules, Options 3, §13(b)(1)(E).

[37] Nasdaq PHLX Rules, Options 3, §13(b)(1)(F). OPRA is the national market system plan responsible for collecting and disseminating consolidated options last-sale and quotation information originating from the participant options exchanges; see SEC, "Options Price Reporting Authority" (OPRA) overview page.

[38] Nasdaq PHLX Rules, Options 3, §13(b)(1)(I).

[39] For the consolidated dissemination role of OPRA, see SEC, OPRA overview; see also SEC orders discussing that transactions executed on options exchanges and options quotations are communicated to the public through OPRA.

[40] Nasdaq PHLX Rules, Options 3, §13(b)(2)(A), §13(b)(2)(B), and §13(b)(2)(F).

[41] Nasdaq PHLX Rules, Options 3, §13(b)(4).

[42] Nasdaq PHLX Rules, Options 3, Supplementary Material .06 to §11 (Facilitation ISO definition); Nasdaq PHLX Rules, Options 3, §13 ("PIXL ISO Order" definition).

NBBO/internal PBBO constraints, and in customer-protection states require improvement over public customer interest at the best prices. In the Facilitation Mechanism, eligible orders must be entered at a price equal to or better than the NBBO and internal PBBO on the same side (and are required to improve when a Public Customer order is at the BBO/internal PBBO), and responses may be submitted at the facilitation price or better.[43] The Solicited Order Mechanism imposes analogous constraints: responses must be entered at prices equal to or better than the better of the internal PBBO or the NBBO, and the mechanism cancels orders if execution would occur at an inferior price to the exchange best bid or offer.[44] These rule-based constraints provide the institutional basis for our "non-ISO" interpretation: auction executions are designed to be no worse than prevailing protected quotations, while competition occurs within a short exposure window and without real-time disclosure of rival responses.

Finally, our SLAN messages are not synonymous with PIXL. SLAN captures single-leg auctions conducted on PHLX that include, but are not limited to, the price improvement auction (PIXL) as well as exposure-based mechanisms such as facilitation and solicitation.[45] PIXL is explicitly structured around a stop price and a price improvement auction format.[46] Other SLAN-eligible mechanisms, however, permit responses at the initiating price or better and impose improvement requirements only in specific states (e.g., when public customer interest is at the best prices).[47] Consequently, even though auction design is highly competitive and often generates improved executions, SLAN data need not exhibit mechanically positive price improvement in every auction message. This distinction motivates our subsequent analysis of the incidence and determinants of price improvement.

[43] Nasdaq PHLX Rules, Options 3, §11(b)(1) and §11(b)(3).
[44] Nasdaq PHLX Rules, Options 3, §11(d)(2) and §11(d)(3)(A).
[45] Nasdaq PHLX Rules, Options 3, §11 (auction mechanisms) and §13 (PIXL).
[46] Nasdaq PHLX Rules, Options 3, §13(b)(1)(A) and §13(b)(1)(G).
[47] Nasdaq PHLX Rules, Options 3, §11(b)(3) and §11(b)(1).

B Execution-Quality Monitoring, Routing Accountability, and Calendar-Month Reporting Boundaries

This appendix explains the institutional channel behind the calendar-month interpretation in Section 3.2. The relevant actors are distinct. A retail customer submits an option order to a broker-dealer. For a non-directed retail order, the broker-dealer generally chooses the routing destination. The destination may be an options exchange, an exchange auction mechanism, an executing broker-dealer, a market center, or a liquidity-provider relationship associated with a wholesaler or affiliated market maker. The liquidity provider or auction responder supplies liquidity and may compete by offering an execution price that improves on the prevailing quote. The broker-dealer is the party with the customer's best-execution obligation; the liquidity provider is not the party that owes that legal duty to the retail customer. Nevertheless, the liquidity provider's auction response helps determine the realized execution price, and that price enters the broker's execution-quality statistics.

FINRA Rule 5310 gives the first institutional link.[48] The rule requires a broker-dealer, when handling a customer order, to use reasonable diligence to identify the best market and obtain a price that is as favorable as possible under prevailing market conditions. The supplementary material also states that a broker-dealer that routes customer orders to other broker-dealers on an automated, non-discretionary basis, or that internalizes customer order flow, must conduct periodic "regular and rigorous" execution quality reviews if it does not conduct order-by-order review. Those reviews must be conducted by security and order type and at least quarterly, although the rule instructs firms to consider whether more frequent reviews are needed for their business. The same supplementary material identifies "price improvement opportunities" as one of the factors that must be considered when comparing current routing and execution arrangements with competing markets. It also lists price disimprovement, likelihood and speed of execution,

[48] FINRA Rule 5310, *Best Execution and Interpositioning*, https://www.finra.org/rules-guidance/rulebooks/finra-rules/5310.

execution size, transaction costs, customer needs, and the existence of internalization or payment-for-order-flow arrangements. Price improvement is therefore not merely a descriptive statistic; it is a stated dimension of the broker's execution-quality review.

Rule 606 of Regulation NMS gives the second institutional link.[49] For non-directed customer orders in NMS securities that are option contracts, the rule requires public order-routing reports at a calendar-quarter frequency, with options routing information reported at calendar-month granularity. Rule 606 also requires disclosure of payment for order flow, profit-sharing payments, transaction fees, and rebates for reportable venues, and it requires discussion of material aspects of routing relationships that may influence routing decisions, including volume-based incentives or order-flow commitments. The reporting party is the broker-dealer, not the wholesaler. The object being reported is the broker's routed order flow. This distinction matters because the channel in the paper does not require the false claim that liquidity providers directly file monthly price-improvement reports with the SEC. The narrower claim is that brokers evaluate and disclose routing performance using calendar-based statistical objects, and price improvement is one dimension of execution quality that can affect future routing decisions.

Enforcement and practitioner evidence are consistent with this allocation of responsibility. In the SEC's 2020 Robinhood order-flow case, the Commission charged a retail broker-dealer with misleading customers about payments from trading firms for routing customer orders and with failing to satisfy its duty to seek the best reasonably available execution terms.[50] The case concerned the broker's customer communications, routing incentives, and execution quality. Public descriptions of retail option routing similarly emphasize that brokers may use exchange auctions to obtain price improvement and that market makers or auction responders supply liquidity in that process.[51]

[49]Rule 606 of Regulation NMS, *Disclosure of Order Routing Information*, 17 CFR 242.606. The SEC's Rule 605/606 page is https://www.sec.gov/about/divisions-offices/division-trading-markets/broker-dealers/disclosure-order-execution-order-routing-information; the current eCFR text is https://www.ecfr.gov/current/title-17/chapter-II/part-242/section-242.606.

[50]SEC Press Release No. 2020-321, Dec. 17, 2020, https://www.sec.gov/newsroom/press-releases/2020-321.

[51]Theresa W. Carey, "Online Broker Battlefield Shifts from Costs to Execution

Combining these pieces gives the calendar-month reporting hypothesis. The broker-dealer is the legally accountable evaluator of execution quality. Routes, venues, market centers, and liquidity-provider relationships are the objects being evaluated. Price improvement is one measured dimension of that quality. Options order-routing disclosures are organized using calendar-month information. A liquidity provider that competes for future retail option flow may therefore have an economic incentive to maintain favorable execution-quality statistics around calendar reporting periods. This interpretation does not imply that every liquidity provider mechanically raises PI at each month-end. It implies a weaker and more plausible prediction: calendar-month position can affect auction concessions when future order-flow allocation depends partly on measured execution quality. Because additional price improvement is costly, any such response should be concentrated where the marginal cost of improving price is low, such as orders with weaker adverse-selection signals, lower inventory risk, or wider quoted spreads that leave more room for economically cheap improvement.

Quality," *Investopedia*, Oct. 25, 2019, https://www.investopedia.com/online-broker-battlefield-shifts-from-costs-to-execution-quality-4774026.

C Auction Aggregation and Conditional Trade Signing

This appendix clarifies why an auction-based setting naturally motivates a conditional trade-signing model. Classic public signing rules infer the initiator side from the transaction price relative to quoted bid and ask prices. In an exchange-run option auction, however, the execution price is the outcome of a short competitive response process. The final price therefore reflects not only the initiating side of the order, but also the state-dependent willingness of liquidity providers to improve upon displayed quotes.

Let r_{ik} denote liquidity provider k's latent response for trade i, expressed as a price-improvement fraction or, more generally, as a response price mapped into the bid–ask spread. A stylized representation of the realized auction response is

$$R_i = \mathscr{A}\{r_{i1}, \ldots, r_{iK_i}\},$$

where K_i is the set of potential responders and $\mathscr{A}(\cdot)$ denotes the auction allocation and pricing rule. Each response may depend on public variables $\mathbf{x}_i$, such as quotes, spreads, depth, contract characteristics, volatility, and short-horizon price dynamics, as well as private variables $\mathbf{z}_{ik}$, such as inventory, risk limits, risk preferences, proprietary signals, and other information not observed by the researcher:

$$r_{ik} = g_k(\mathbf{x}_i, \mathbf{z}_{ik}).$$

The econometrician does not observe the full response book or the private state variables. The empirical object is therefore not the structural response rule $g_k(\cdot)$, but the reduced-form conditional probability of the exchange-recorded initiator side given public information:

$$\Pr(q_i = 1 \mid \mathbf{x}_i).$$

This distinction is important for interpreting the use of XGBoost. A gradient-boosted tree

classifier is well suited to this setting because it can approximate threshold-dependent and interaction-rich response patterns in the public component of the auction state. Conditional on a trained ensemble, the fitted logit can be written as an additive combination of tree-indicator functions:

$$F(\mathbf{x}_i) = c + \sum_{m=1}^{M} \sum_{\ell \in \mathscr{L}_m} w_{\ell m} \mathbf{1}\{\mathbf{x}_i \in R_{\ell m}\}, \qquad \widehat{p}_i = \Lambda\{F(\mathbf{x}_i)\},$$

where $\mathscr{L}_m$ denotes the leaves of tree m, $R_{\ell m}$ is the feature-space region associated with leaf ℓ, $w_{\ell m}$ is the corresponding leaf score, and $\Lambda(z) = 1/(1+\exp(-z))$. The final classifier applies a threshold to this fitted probability:

$$\widehat{q}_i = \mathbf{1}\{\widehat{p}_i \geq \tau\}.$$

Thus, the model can be represented as a thresholded additive combination of tree-indicator rules over the public information set.

This representation should not be interpreted as a claim that liquidity providers literally use the fitted decision trees. The trees are a flexible reduced-form approximation to the public component of a richer auction mechanism. The drawback is that the researcher does not observe all variables entering liquidity providers' true decisions, especially private inventory and risk-management information. The empirical premise is instead that public market variables form the main observable state variables available for trade signing, and that their nonlinear interactions contain economically meaningful information about the latent initiator side. The SHAP and LASSO layers then discipline this flexible approximation by identifying which public signals have incremental explanatory content in out-of-sample trade-direction recovery.

D Additive Tree Approximation of Auction Outcomes

This section formalizes the auction mechanism behind our tree-based classifier. We define an indicator for whether the best liquidity-provider response provides more than 50% price improvement and therefore crosses the midpoint. We then show that if individual automated responses follow tree-based threshold rules, the auction-level outcome can be viewed as an aggregation of these rules, motivating an additive tree approximation. Finally, we combine this auction outcome with the midpoint rule to infer trade direction.

For auction i, let B_i and A_i denote the prevailing quoted bid and ask, $S_i = A_i - B_i > 0$ the quoted spread, and $m_i = (A_i + B_i)/2$ the quote midpoint. Let $D_i \in \{+1, -1\}$ denote the economically correct trade direction, where $D_i = +1$ indicates a buyer-initiated trade and $D_i = -1$ indicates a seller-initiated trade. Each liquidity provider $j \in \mathscr{J}_i$ submits a response price r_{ij}. Conditional on the initiating direction, the normalized price improvement offered by liquidity provider j is

$$PI_{ij} = \begin{cases} \dfrac{A_i - r_{ij}}{S_i}, & \text{if } D_i = +1, \\ \dfrac{r_{ij} - B_i}{S_i}, & \text{if } D_i = -1. \end{cases}$$

This normalization measures price improvement as a fraction of the quoted spread. Thus, $PI_{ij} = 0$ corresponds to no improvement relative to the relevant quoted side, while $PI_{ij} = 1/2$ corresponds to a response at the midpoint. The key threshold for trade classification is therefore whether the response improves the price by more than one half of the quoted spread. Define

$$Z_{ij} = \mathbf{1}\{PI_{ij} > 1/2\}.$$

Here $Z_{ij} = 1$ indicates that liquidity provider j's response crosses the midpoint from the perspective of the initiating order. Since the auction selects the best available response, the auction-level price

improvement is

$$PI_i = \max_{j \in \mathscr{J}_i} PI_{ij},$$

and the corresponding auction-level crossing indicator is

$$Z_i = \mathbf{1}\{PI_i > 1/2\} = \max_{j \in \mathscr{J}_i} Z_{ij}.$$

Thus, $Z_i = 1$ if at least one participating liquidity provider offers more than 50% price improvement, and $Z_i = 0$ otherwise.

This structure provides a direct motivation for using a tree-based classifier. Auction responses are submitted within a very short time window, so liquidity providers' decisions are likely to be governed by automated rules based on observable auction and market conditions. Let X_i denote these covariates. Suppose liquidity provider j's normalized price improvement can be represented by a decision tree,

$$PI_{ij} = g_j(X_i) = \sum_{\ell=1}^{L_j} c_{j\ell} \mathbf{1}\{X_i \in R_{j\ell}\},$$

where $\{R_{j\ell}\}_{\ell=1}^{L_j}$ are the terminal-node regions of liquidity provider j's decision tree and $c_{j\ell}$ is the price improvement associated with terminal node ℓ. The binary response decision can then be written as

$$Z_{ij} = \mathbf{1}\{g_j(X_i) > 1/2\} = \sum_{\ell=1}^{L_j} \mathbf{1}\{c_{j\ell} > 1/2\} \mathbf{1}\{X_i \in R_{j\ell}\}.$$

Hence, Z_{ij} is itself a threshold rule generated by a decision tree. The auction outcome aggregates these individual response rules across participating liquidity providers:

$$Z_i = \max_{j \in \mathscr{J}_i} Z_{ij} = 1 - \prod_{j \in \mathscr{J}_i} (1 - Z_{ij}).$$

This expression shows that the auction outcome is an "OR" rule over liquidity providers' automated response decisions: the auction crosses the midpoint if any participating liquidity provider submits a sufficiently aggressive response.

In practice, the econometrician does not observe each liquidity provider's full response rule $g_j(\cdot)$, and the set of participating liquidity providers may vary across auctions. Therefore, rather than modeling each Z_{ij} separately, we approximate the auction-level conditional probability by an additive ensemble of trees:

$$\Pr(Z_i = 1 \mid X_i) = \Lambda\left(F(X_i)\right), \qquad F(X_i) = \sum_{k=1}^{K} \nu f_k(X_i),$$

where $\Lambda(u) = 1/(1+\exp(-u))$, $f_k(\cdot)$ is a regression tree, and ν is the learning rate. This additive representation is consistent with the auction mechanism. Individual trees capture threshold-based automated response rules, while the ensemble approximates the auction-level aggregation of many such rules. Sequential boosting improves this approximation by adding trees that correct errors from previous trees, allowing the model to capture nonlinearities, threshold effects, and interactions in liquidity providers' responses.

Finally, the auction-level crossing indicator Z_i determines whether the midpoint rule must be reversed. Let $\widetilde{D}_i$ denote the direction implied by the midpoint rule:

$$\widetilde{D}_i = \begin{cases} +1, & \text{if } P_i > m_i, \\ -1, & \text{if } P_i < m_i, \end{cases}$$

where P_i is the auction execution price. Abstracting from the knife-edge case $P_i = m_i$, if $Z_i = 0$, the execution price remains on the same side of the midpoint as the initiating order, so the midpoint rule gives the correct direction. If $Z_i = 1$, the execution price crosses the midpoint because the auction provides more than 50% price improvement, so the midpoint-rule classification must be reversed. Therefore,

$$D_i = (1 - 2Z_i)\widetilde{D}_i.$$

Equivalently, if $Y_i = \mathbf{1}\{D_i = +1\}$ is the buyer-initiated indicator and $\widetilde{Y}_i = \mathbf{1}\{P_i > m_i\}$ is the

midpoint-rule buyer indicator, then

$$Y_i = \widetilde{Y}_i(1 - Z_i) + (1 - \widetilde{Y}_i)Z_i.$$

This decomposition provides a stylized motivation for conditional trade signing. Away from the midpoint, predicting whether price improvement crosses the midpoint is equivalent to determining whether the midpoint-implied sign should be retained or reversed. At the midpoint, however, price location is uninformative and trade direction must be inferred directly from the broader market state. Accordingly, our empirical implementation does not estimate Z_i as a separate supervised target. Instead, it estimates $\Pr(q_i = 1 \mid X_i)$ directly within the five observable execution-price location groups described in Section 2.2.

E Machine Learning for Trade-Direction Identification

This subsection develops our trade-direction identification framework. Because execution price alone is insufficient to recover the initiator side in option auctions, we replace midpoint-based signing with a data-driven procedure that exploits the set of public market information surrounding each trade. The approach combines flexible nonlinear modeling with disciplined feature screening to infer trade direction in a setting where the mapping from price to initiator is inherently state dependent.

E.1 Identification Setup

For each completed trade i, let $q_i \in \{0,1\}$ denote the initiator side defined in (1). This direction is recorded by the exchange but is not included in public data feeds. Instead, the econometrician observes a vector of public market information, $\mathbf{x}_i \in \mathbb{R}^p$, constructed from option quotes and trades, the underlying asset, and related market-state variables surrounding the trade.

Our objective is to recover the realized trade direction q_i using $\mathbf{x}_i$. Following standard trade-classification approaches, the information set may include variables observed at or after execution, including short-horizon post-trade public signals, consistent with the inherently ex post nature of trade signing. As shown in Section 2.1, execution price alone is no longer sufficient for this purpose in option auctions, because the relation between price location and initiator side is not stable across trades. More generally, even a rich public information set does not fully reveal the initiator side, as auction outcomes depend on competitive responses and other factors not directly observable in public data.

The relevant object is therefore the conditional probability $\Pr(q_i = 1 \mid \mathbf{x}_i)$, which captures how the likelihood of buyer initiation varies with the surrounding market state. The identification problem is to infer the already-realized trade direction from the available public information, rather than to apply a deterministic signing rule based on execution price alone.

E.2 GS-LASSO Identification Procedure

The identification problem above requires a method that can accommodate nonlinearities and interactions across a high-dimensional set of public signals, while retaining a disciplined and reproducible feature-selection rule. A suitable procedure must therefore satisfy two requirements simultaneously. First, it must flexibly approximate the conditional mapping from $\mathbf{x}_i$ to q_i, allowing for interactions and threshold effects across market variables. Second, it must provide a disciplined way to isolate a stable and interpretable subset of predictors from a large candidate set. We therefore develop a procedure that combines flexible nonlinear modeling with structured screening. We refer to this approach as the *GS-LASSO* framework.

Nonlinear representation via gradient boosting. We begin by estimating a gradient-boosted tree classifier (XGBoost) on the full set of candidate predictors within a training window. Let $\mathscr{W}$ denote a training window consisting of a contiguous block of trades over a fixed calendar period. All model estimation and feature selection steps described below are performed using only observations in $\mathscr{W}$. Let $\hat{p}_i^{(1)}$ denote the fitted probability that $q_i = 1$, obtained from the gradient-boosted tree ensemble using predictors $\mathbf{x}_i$. The first-stage estimator fits a flexible nonlinear classifier that approximates the conditional probability $\Pr(q_i = 1 \mid \mathbf{x}_i)$ using an ensemble of decision trees.[52] Formally, it solves

$$\mathscr{L}^{(1)}(\mathscr{W}) = \sum_{i \in \mathscr{W}} \ell\left(q_i, \hat{p}_i^{(1)}\right) + \sum_{m=1}^{M} \Omega(f_m), \tag{6}$$

[52] This reduced-form approximation has a direct auction interpretation. Liquidity providers' response rules may depend on public state variables as well as private variables such as inventory, risk limits, risk preferences, and proprietary signals. XGBoost is useful because tree ensembles can approximate threshold-dependent response rules and interactions over the public component of that state (Chen and Guestrin, 2016), and related finance applications use machine learning to recover latent trading-related objects (Bogousslavsky et al., 2024; Easley et al., 2021). The limitation is that the model is not a structural model of providers' decision trees: omitted private information means that the classifier estimates $\Pr(q_i = 1 \mid \mathbf{x}_i)$ using the public information set rather than the full auction state. Online Appendix C formalizes this interpretation.

where $\ell(\cdot)$ is the logistic loss and each f_m denotes a tree-based component of the ensemble. The regularization term is

$$\Omega(f_m) = \gamma T_m + \frac{1}{2}\lambda_{\text{XGB}} \sum_{\ell=1}^{T_m} w_{m\ell}^2,$$

where T_m indexes the complexity of tree m (number of leaves) and $w_{m\ell}$ are the associated node weights. The penalty parameters γ and λ_{XGB} control model complexity and prevent overfitting. As a tree-based ensemble, this estimator partitions the state space and captures nonlinearities and high-order interactions without imposing a parametric structure. This stage therefore provides a flexible approximation to the conditional probability $\Pr(q_i = 1 \mid \mathbf{x}_i)$ using the full public information set.

Additive representation via SHAP. The fitted nonlinear model is then mapped into an additive representation using SHapley Additive exPlanations (SHAP values) computed within $\mathscr{W}$. Following Lundberg and Lee (2017), the fitted log-odds can be decomposed as

$$\text{logit}\left(\hat{p}_i^{(1)}\right) = \phi_0 + \sum_{j=1}^{p} \phi_{ij}, \tag{7}$$

where ϕ_{ij} represents the contribution of feature j to the prediction for trade i. This step provides an additive decomposition of the *fitted nonlinear prediction*, placing feature contributions on a common scale that facilitates subsequent LASSO-based screening. As a result, it yields a representation in which feature contributions can be compared and screened in a coherent way.

Sparse screening via LASSO on SHAP values. Using the SHAP-transformed representation, we next perform feature screening via logistic LASSO. Specifically, we estimate

$$(\hat{\beta}_0, \hat{\beta}) = \arg\min_{\beta_0,\beta} \left\{ -\frac{1}{|\mathscr{W}|} \sum_{i\in\mathscr{W}} \left[q_i(\beta_0 + \phi_i'\beta) - \log\left(1 + e^{\beta_0 + \phi_i'\beta}\right) \right] + \lambda_{\text{LASSO}} \sum_{j=1}^{p} |\beta_j| \right\}, \tag{8}$$

where $\phi_i = (\phi_{i1}, \ldots, \phi_{ip})'$, and λ_{LASSO} is selected via cross-validation within $\mathscr{W}$. The selected feature set is

$$\widehat{\mathscr{S}} = \left\{ j \in \{1, \ldots, p\} : \hat{\beta}_j \neq 0 \right\}.$$

This stage serves as a structured screening device that builds directly on the SHAP representation. SHAP maps the fitted nonlinear model into an additive decomposition in log-odds space, placing feature contributions on a common scale suitable for selection. The LASSO step then operates on this representation to perform *joint* feature selection, retaining a sparse subset of predictors that provide incremental explanatory power conditional on the others. In this sense, SHAP provides a nonlinear-to-additive mapping, while LASSO imposes disciplined sparsity on that mapping. The resulting set $\widehat{\mathscr{S}}$ therefore consists of predictors whose contributions remain relevant for trade-direction identification after accounting for the nonlinear structure learned in the first-stage model, rather than features that appear important only in isolation.

Nonlinear re-estimation on the selected set. The final step returns to nonlinear modeling. We re-estimate XGBoost on the same training window $\mathscr{W}$ using only the selected raw predictors $\mathbf{x}_{i,\widehat{\mathscr{S}}}$, producing fitted probabilities $\hat{p}_i^{(2)}$. The final classification rule is

$$\hat{q}_i = \mathbf{1}\left\{ \hat{p}_i^{(2)} \geq 0.5 \right\}.$$

This step is essential because the role of the LASSO stage is feature selection rather than final prediction. Once the relevant predictors have been identified, re-estimating a flexible nonlinear model allows the procedure to retain interaction effects and threshold behavior within a reduced and more stable feature space.

Summary of the procedure. Within a given training window $\mathscr{W}$, the GS-LASSO framework proceeds as follows:

1. estimate a first-stage XGBoost model on the full feature set;

2. compute SHAP values to obtain an additive representation of the fitted model;

3. apply cross-validated logistic LASSO on the SHAP values to select features;

4. re-estimate XGBoost on the selected raw predictors and construct the final classifier.

All steps are confined to the training window $\mathscr{W}$, ensuring that feature selection, model estimation, and hyperparameter choices do not use information from the evaluation period.

E.3 Rolling Design and Evaluation

In the main empirical implementation, we apply the GS-LASSO procedure separately within five observable price-location routes. The route partition is determined only by public execution-price and quote information: trades strictly inside the bid–ask spread are split according to whether the transaction price is below, at, or above the quote midpoint, while non-inside trades are split according to whether the midpoint rule signs them as seller- or buyer-initiated. This gives five mutually exclusive groups: at ask, at bid, exact midpoint, inside below midpoint, inside above midpoint. This partition is not based on the true trade direction or on realized price improvement. It is motivated by the midpoint-failure mechanism itself: the same public price location represents different identification problems depending on whether the trade is at an endpoint, strictly inside the spread, or exactly at the midpoint. Estimating route-specific models therefore allows the selected features, nonlinear interactions, and validation thresholds to differ across economically distinct signing problems rather than forcing one global mapping from price location to trade direction.

We implement the GS-LASSO procedure in a rolling-window design. Let t index calendar time at the level of non-overlapping periods (e.g., weeks). To estimate trade directions in period $t+1$, we use observations from period t as the training sample. Specifically, let $\mathscr{W}_t$ denote the set of trades in period t. All model estimation, SHAP decomposition, and feature-selection steps are performed within $\mathscr{W}_t$, and the resulting classifier is applied to trades in $\mathscr{W}_{t+1}$. This rolling structure allows

the mapping from public information to trade direction to evolve over time, while maintaining sufficient sample size within each window for stable estimation. We evaluate performance using out-of-sample classification accuracy:

$$\text{Accuracy} = \frac{1}{N_{\mathscr{W}_{t+1}}} \sum_{i \in \mathscr{W}_{t+1}} \mathbf{1}\{\hat{q}_i = q_i\}, \tag{9}$$

where $N_{\mathscr{W}_{t+1}}$ is the total number of trades in $\mathscr{W}_{t+1}$.

This design ensures a strict separation between model construction and evaluation, so that all reported performance reflects genuine out-of-sample identification. We benchmark the GS-LASSO procedure against the midpoint rule on the same set of trades to assess whether a richer public information set can more accurately recover the exchange-recorded trade direction in an auction environment where execution price alone is no longer sufficient.[53]

[53] We also examine alternative implementations, including static estimation windows and a plain XGBoost classifier without SHAP-LASSO screening. These exercises assess whether the gains in trade-direction identification arise from the rolling design, from nonlinear modeling alone, or from the combination of nonlinear representation and structured feature selection.

F GS-LASSO Simulation Design and Results

To evaluate whether a low-capacity gradient-boosted tree representation can provide a useful screening step before fitting a high-capacity XGBoost model, we conduct a 100-repetition teacher-student simulation. The simulation is designed to stress the learner in three ways: the outcome is generated by a nonlinear tree-based mechanism using only true predictors, the observed design also contains many false predictors, and the false predictors have a training-test variance shift that makes them easier to overfit in the training sample.

In each Monte Carlo repetition $r = 1, \ldots, 100$, we set different random seed to generate 100 true features and 1,000 false features. Let

$$X_i = (X'_{i,T}, X'_{i,F})',$$

where $X_{i,T} \in \mathbb{R}^{100}$ denotes the true-feature block and $X_{i,F} \in \mathbb{R}^{1000}$ denotes the false-feature block. For the true features, all rows are drawn from $N(0, I_{100})$. For the false features, the first 5,000 rows are drawn with variance $4I_{1000}$, while the remaining 95,000 rows are drawn with variance I_{1000}. Thus, the false features are not causal, but their marginal distribution is deliberately less stable across the training and test samples.

The teacher model is built using only the 100 true features. First, we draw a separate teacher-training sample and generate auxiliary binary labels from a nonlinear rule containing 24 main effects, 36 pairwise interactions, and 96 threshold clauses with depths between two and four. These auxiliary labels are used only to train the teacher. We then train an XGBoost classifier with 300 trees, maximum depth five, learning rate 0.05, subsampling rate 0.9, and column-sampling rate 0.8. The saved XGBoost teacher, not the auxiliary rule itself, defines the base data-generating process.

The released simulation sample contains 100,000 observations. For each observation, the saved teacher predicts a base probability using only $X_{i,T}$. We then add one rare but repeatable true-feature

pattern. This pattern uses the first five true features and is activated when all five exceed $\Phi^{-1}(0.45)$, giving an approximate prevalence of $(0.55)^5 \approx 5.03\%$. For observations satisfying this pattern, we add 10 to the teacher logit before converting probabilities into binary labels. Formally, if p_i^0 is the teacher probability and R_i indicates the rare-pattern event, the clean label is generated from

$$p_i = \Lambda\left\{\log\left(\frac{p_i^0}{1-p_i^0}\right) + 10R_i\right\}, \qquad Y_i^{\text{clean}} = \mathbf{1}\{p_i \geq 0.5\}.$$

Finally, to introduce training-label noise, we randomly flip 5% of the labels among the first 5,000 rows. The test labels are not flipped.

Student models are trained on a fixed row-order split. The first 5,000 observations form the training set, and the remaining 95,000 observations form the test set. We estimate two baseline XGBoost students using learning rate 0.05, subsampling rate 1, column-sampling rate 1, minimum child weight 1, zero split penalty, and zero ℓ_1 and ℓ_2 regularization. The small model uses 100 trees with maximum depth four. The large model uses 2,000 trees with maximum depth 25.

We then implement two SHAP/LASSO screening procedures. In the first, SHAP values are computed from the large XGBoost model; in the second, SHAP values are computed from the small XGBoost model. In both cases, we compute the XGBoost SHAP contribution matrix on the 5,000-row training sample, standardize the SHAP columns, and estimate a logistic LASSO using three-fold stratified cross-validation, a 20-point penalty grid, and a two-standard-error selection rule. The selected original features are then passed to a refitted 2,000-tree, depth-25 XGBoost model.

Table F1 reports means and standard deviation across the 100 repetitions. The small model uses, on average, 58.88 true features and 331.38 false features. The unrestricted large model uses all 100 true features and all 1,000 false features, fits the noisy training set perfectly, and achieves 80.75% test accuracy. Applying SHAP/LASSO to the large model selects all 1,100 features, so it reproduces the unrestricted large-model result. By contrast, applying SHAP/LASSO to the small model selects a much smaller feature set: 56.09 true features and 288.29 false features on average.

Refitting the large XGBoost model on this selected set yields the highest average test accuracy, 81.55%, and the highest rare-pattern test accuracy, 64.07%. Thus, the small-model SHAP/LASSO screen removes about 711.71 false features relative to the unrestricted large model and improves test accuracy by about 0.80% with a t-statistic of 37.39.

Table F1 GS-LASSO Simulation Results Across 100 Repetitions

This table reports Monte Carlo means across 100 simulation repetitions. Standard deviations across repetitions are reported in parentheses. In each repetition, the data contain 100 true features and 1,000 false features. Labels are generated by a saved XGBoost teacher model using only the true features, with an additional rare true-feature pattern and 5% label noise in the training sample. The first 5,000 observations are used for training and the remaining 95,000 observations are used for testing. "True Used" and "False Used" report the number of true and false features used by the fitted XGBoost model. "Rare Test Acc." reports test accuracy on observations satisfying the rare-pattern condition. Small Model uses 100 trees with depth 4. Large Model uses 2,000 trees with depth 25. Large Model + LASSO computes SHAP values from the Large Model, applies logistic LASSO screening, and refits the Large Model using the selected features. GS-LASSO computes SHAP values from the Small Model, applies logistic LASSO screening, and refits the Large Model using the selected features.

Model	Train Acc.	Test Acc.	True Used	False Used	Rare Test Acc.
Small Model	0.9072	0.8049	58.88	331.38	0.6094
	(0.0059)	(0.0119)	(4.9935)	(17.3070)	(0.0684)
Large Model	1.0000	0.8075	100.00	1000.00	0.6264
	(0.0000)	(0.0108)	(0.0000)	(0.0000)	(0.0715)
Large Model + LASSO	1.0000	0.8075	100.00	1000.00	0.6264
	(0.0000)	(0.0108)	(0.0000)	(0.0000)	(0.0715)
GS-LASSO	**1.0000**	**0.8155**	**56.09**	**288.29**	**0.6407**
	(0.0000)	**(0.0105)**	**(4.6971)**	**(13.8850)**	**(0.0692)**

G Additional Evidence on Quote Mismatching

G.1 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, Battalio et al. (2026) show that adjusting for quote mismatches significantly improves the accuracy of the Lee and Ready (1991) trade classification algorithm in the stock market relative to its baseline accuracy of 86%. 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.[54] 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 G1 presents the time series of timestamp differences between OPRA and PHLX. On 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,

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

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.[55] 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. Because OPRA timestamps are recorded to the nearest millisecond, these sub-millisecond differences are too small to explain midpoint-rule accuracy near 55%. We therefore expect timestamp alignment to have little effect on our classification results.

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 G1 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 average daily midpoint-rule accuracy ranges from 50.41% to 55.34% across the alternative quote timestamps. The highest accuracy is obtained using the contemporaneous quote (55.34%), followed by the next quote after the trade (55.32%) and the last quote before the trade (55.10%). Thus, using an alternative quote timestamp does not improve on the contemporaneous midpoint rule. These figures are substantially lower than the accuracy 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

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

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 lower midpoint rule accuracy, further rejecting the quote mismatch explanation.[56]

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.

[56]Table 5 shows that the slower quote updates group (above median) achieves a 53.81% accuracy for the midpoint rule, whereas the faster quote updates group (below median) achieves 56.81%.

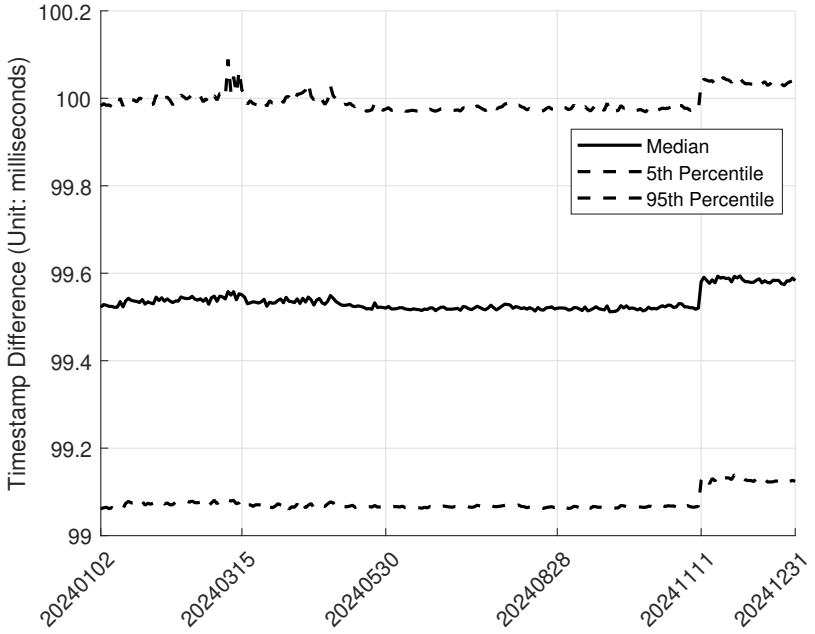
Table G1 Midpoint Rule Accuracy Using Alternative Quote Timestamps

This table reports summary statistics of daily midpoint-rule accuracy using option NBBO quotes sampled at alternative timestamps. For each quote timestamp, the midpoint rule classifies a trade as buyer-initiated when its execution price is above the corresponding quote midpoint and seller-initiated when it is below the quote midpoint. Trades executed exactly at the quote midpoint receive no directional assignment and are counted as incorrect. The contemporaneous-quote specification uses the same midpoint-direction measure reported in Table 2. The mean, standard deviation, minimum, and maximum are calculated across the 252 trading days in 2024. The sample includes all single-leg auction, non-ISO (SLAN) option trades written on stocks and ETFs from January 1, 2024 to December 31, 2024.

Quote timestamp	Mean	Std	Min	Max
Panel A. Stock and ETF options				
1 second before trade	50.41%	2.13%	41.91%	55.18%
100 milliseconds before trade	54.75%	2.45%	44.65%	59.94%
Last quote before trade	55.10%	2.40%	45.47%	60.19%
Contemporaneous quote	55.34%	2.39%	45.85%	60.47%
Next quote after trade	55.32%	2.36%	45.85%	60.35%
1 second after trade	54.26%	2.01%	46.73%	59.21%
Panel B. Stock options				
1 second before trade	51.13%	2.02%	44.89%	55.79%
100 milliseconds before trade	55.59%	2.17%	48.86%	60.71%
Last quote before trade	55.77%	2.15%	49.16%	60.92%
Contemporaneous quote	56.00%	2.11%	49.56%	61.05%
Next quote after trade	55.93%	2.10%	49.50%	61.00%
1 second after trade	54.88%	1.86%	49.27%	60.01%
Panel C. ETF options				
1 second before trade	47.08%	3.73%	34.01%	54.48%
100 milliseconds before trade	50.83%	5.41%	34.92%	65.81%
Last quote before trade	51.91%	5.35%	36.65%	66.19%
Contemporaneous quote	52.21%	5.34%	37.08%	66.35%
Next quote after trade	52.35%	5.31%	37.29%	66.34%
1 second after trade	51.37%	3.96%	39.80%	59.32%

Figure G1. PHLX Auction Duration in Matched 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.



H Low-Complexity All-Feature XGBoost Benchmark

This appendix evaluates the incremental contribution of the complete GS-LASSO pipeline relative to its low-complexity first-stage model. The benchmark estimates five price-location-specific XGBoost classifiers using all 445 candidate features, 200 boosting rounds, a maximum tree depth of six, a learning rate of 0.05, row and feature subsampling rates of 0.8, an L_2 penalty of 10, and an L_1 penalty of 0.1. Predictions use the standard 0.5 threshold. The benchmark does not apply SHAP-LASSO screening, the second-stage high-complexity parameter search, or the subsequent deeper refit. The comparison therefore asks whether feature screening followed by a higher-complexity re-estimation adds predictive value beyond the regularized nonlinear representation learned in the first stage.

We implement the benchmark using the same one-day, one-week, and one-month training horizons and the same static and rolling updating schemes used in the model-specification analysis. Within each specification, Low-Complexity XGBoost and GS-LASSO are evaluated on exactly the same trading dates. The reported difference is the mean paired daily accuracy difference, defined as GS-LASSO accuracy minus Low-Complexity XGBoost accuracy.

The incremental value of the full pipeline increases with the length of the training window. Under the one-month rolling specification, mean daily accuracy rises from 79.23% for Low-Complexity XGBoost to 80.46% for GS-LASSO, a gain of 1.23 percentage points. The corresponding gains are 0.75 percentage points for the one-month static specification and 0.73 percentage points for the one-week rolling specification. The gains are smaller for the one-day rolling and one-week static specifications, at 0.19 and 0.18 percentage points, respectively. Under the one-day static specification, the models perform economically almost identically, with GS-LASSO lower by 0.04 percentage points. This pattern is consistent with the SHAP-LASSO screen and high-complexity refit becoming most useful when the training window contains sufficient information to identify stable predictors and more complex interactions.

Table H1 GS-LASSO and the Low-Complexity All-Feature XGBoost Benchmark

This table compares daily out-of-sample classification accuracy for the complete GS-LASSO pipeline and a low-complexity first-stage XGBoost benchmark using all 445 candidate features. The benchmark uses five price-location-specific classifiers, 200 boosting rounds, maximum depth six, learning rate 0.05, subsample and column-subsample rates of 0.8, L_2 penalty 10, L_1 penalty 0.1, and the standard 0.5 classification threshold. It omits SHAP-LASSO screening and the second-stage high-complexity refit. For each specification, the two models are evaluated on the same trading dates. Difference is GS-LASSO minus Low-Complexity XGBoost in percentage points. The t-statistic is based on the time series of paired daily accuracy differences and tests whether their mean equals zero. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Specification	Obs	Low-Complexity XGBoost	GS-LASSO	Difference	Paired t-statistic
1 Day Static	251	77.58%	77.55%	-0.04***	-4.83
1 Day Rolling	251	78.83%	79.02%	0.19***	17.85
1 Week Static	247	78.32%	78.50%	0.18***	23.53
1 Week Rolling	247	79.32%	80.05%	0.73***	57.45
1 Month Static	231	78.54%	79.28%	0.75***	65.27
1 Month Rolling	231	79.23%	80.46%	1.23***	79.81

I Price Improvement and Trade Informativeness

This appendix provides descriptive evidence on the relation between price improvement and public signals associated with trade direction. The regressions are not intended to identify causal determinants of price improvement, recover liquidity providers' proprietary pricing rules, or measure their private information and inventory positions. Instead, they examine whether price improvement covaries with the same observable market state that helps GS-LASSO recover trade direction.

We first construct a model-implied measure of public directional signal strength using predictions from the current one-month rolling GS-LASSO specification. Let $\widehat{p}_{i,t}$ denote the predicted probability that trade i at time t is buyer-initiated. We define

$$\widehat{Info}_{i,t} = |\widehat{p}_{i,t} - 0.5| .$$

Values close to zero correspond to trades for which the public signals provide weak directional information, while larger values correspond to trades with stronger model-implied directional signals. We estimate

$$\mathrm{PI}_{i,t} = \beta_0 + \beta_1 \widehat{Info}_{i,t} + \alpha_{r(i)} + \gamma_t + \varepsilon_{i,t}, \tag{I1}$$

where $r(i)$ denotes the option-root ticker. Thus, $\alpha_{r(i)}$ represents option-root ticker fixed effects based on the Root field, such as AAPL, SPY, and QQQ, and γ_t represents trading-date fixed effects.

Table I1 Price Improvement and Model-Implied Trade Informativeness

This table reports descriptive regressions of price improvement on model-implied trade informativeness. The dependent variable is price improvement, defined in Equation (1). Model-implied informativeness is $\widehat{Info}_i = |\widehat{p}_i - 0.5|$, where $\widehat{p}_i$ is the predicted probability of buyer initiation from the current one-month rolling GS-LASSO specification. Columns differ in the inclusion of option-root ticker and trading-date fixed effects. Option-root ticker fixed effects are based on the Root field, such as AAPL, SPY, and QQQ; they are not option-contract or option-series fixed effects. Date fixed effects are trading-date fixed effects. The sample includes out-of-sample single-leg auction, non-ISO (SLAN) option trades written on stocks and ETFs from February 1, 2024 to December 31, 2024. Standard errors, clustered by trading date, are reported in parentheses below the coefficient estimates. The regressions describe conditional associations and do not identify causal effects. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

	Dependent Variable: PI		
	(1)	(2)	(3)
$\widehat{Info}$	-0.2161***	-0.2122***	-0.2007***
	(0.0073)	(0.0073)	(0.0079)
Constant	0.5129***	0.5114***	0.5072***
	(0.0012)	(0.0012)	(0.0012)
Option-Root Ticker Fixed Effects	Yes	Yes	No
Trading-Date Fixed Effects	Yes	No	Yes
Adj. R^2	0.0244	0.0210	0.0126
Obs	43,308,025	43,308,025	43,308,025

Table I1 reports a negative and statistically significant coefficient on $\widehat{Info}$ in all three fixed-effect specifications. The association therefore remains when option-root ticker fixed effects are excluded; it is only modestly attenuated from -0.2161 to -0.2007. A 0.10 increase in model-implied directional signal strength is associated with approximately 2.0 to 2.2 percentage points less price improvement. This is a descriptive conditional association rather than evidence that model confidence causally determines execution pricing.

We next examine percentage price improvement in relation to the ten leading public signals in Figures 5 and 6. The dependent variable in Table I2 is measured in percentage points, and each continuous explanatory variable is standardized before estimation. Consequently, each

coefficient describes the percentage-point association with a one-standard-deviation increase in the corresponding signal, conditional on the other included signals and fixed effects. The Table I1 sample is the February–December out-of-sample prediction sample, whereas Table I2 uses the full-calendar-year complete-case sample. The coefficient scales should therefore not be compared mechanically across the two tables.

Price improvement therefore covaries with market conditions relevant for trade-direction recovery. These descriptive associations are consistent with state-dependent execution concessions, but do not identify causal mechanisms or proprietary information.

Table I2 Price Improvement and Leading GS-LASSO Signals

This table reports descriptive fixed-effects regressions of percentage price improvement on the ten leading signals identified by the GS-LASSO feature-importance analysis. The dependent variable is price improvement in percentage points, defined as 100× price improvement. Continuous explanatory variables are standardized to have mean zero and unit standard deviation before estimation. Columns differ in the inclusion of option-root ticker and trading-date fixed effects. Option-root ticker fixed effects are based on the Root field, such as AAPL, SPY, and QQQ; they are not option-contract or option-series fixed effects. Date fixed effects are trading-date fixed effects. The sample includes single-leg auction, non-ISO (SLAN) option trades written on stocks and ETFs during calendar year 2024 for which the dependent variable and all ten explanatory variables are nonmissing. Feature definitions are provided in Online Appendix M. Standard errors, clustered by trading date, are reported in parentheses. The regressions describe conditional associations and do not identify causal effects or proprietary pricing rules. Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

	Dependent Variable: PI%		
	(1)	(2)	(3)
Quoted Spread ($t-100$ms)	-1.4625***	-1.4333***	-1.4588***
	(0.1277)	(0.1283)	(0.1256)
Trade Size	-1.9688***	-1.9010***	-1.9739***
	(0.2187)	(0.2106)	(0.2194)
Displayed NBBO Depth Ratio	0.7307***	0.7198***	0.7303***
	(0.0198)	(0.0235)	(0.0198)
Option Return ($t-1$s, t)	-0.7527***	-0.7664***	-0.7526***
	(0.0098)	(0.0104)	(0.0097)
Surface Price Gap	0.0314*	0.0313*	0.0263
	(0.0175)	(0.0182)	(0.0178)
Underlying VWAP Gap (5m)	0.0494***	0.0619***	0.0455***
	(0.0104)	(0.0118)	(0.0099)
Price Position within NBBO	-0.5368***	-0.5580***	-0.5419***
	(0.0689)	(0.0709)	(0.0691)
Option Return ($t-100$ms, t)	-0.0444**	-0.0348	-0.0458**
	(0.0207)	(0.0215)	(0.0208)
Delta-Neutral PnL ($t+1$m)	-0.0416*	-0.0461*	-0.0448*
	(0.0236)	(0.0249)	(0.0249)
Quoted Spread (t)	2.2234***	2.1923***	2.2231***
	(0.1275)	(0.1273)	(0.1281)
Option-Root Ticker Fixed Effects	Yes	No	Yes
Trading-Date Fixed Effects	Yes	Yes	No
Adj. R^2	0.0207	0.0097	0.0174
Obs	45,929,520	45,929,520	45,929,520

J Trade-Direction Measurement and SLAN Gross Performance

This appendix provides construction details for the gross-performance analysis in Section 3.5.3 and extends the direction-measure comparison in Table 8 with estimates under the baseline GS-LASSO-implied direction and the fixed public-data decision trees. It also reports a complete-sample robustness exercise for exchange-recorded and GS-LASSO-implied direction.

Following Bryzgalova et al. (2023), define the gross performance of trade j, initiated on day t, under direction input k and holding horizon h as

$$\text{GPerf}_{j,t}^{k}(h) = 100 D_{j,t}^{k} \widetilde{Q}_{j,t} \left(V_{j,t+h} - M_{j,t} \right), \tag{10}$$

where $D_{j,t}^{k} \in \{-1, 1\}$ is the signed trade direction under method k, and $\widetilde{Q}_{j,t} = \min\{Q_{j,t}, c_t\}$ is trade size capped at the 99.5th percentile c_t of trade size on day t, following Bryzgalova et al. (2023). $M_{j,t}$ is the option midpoint at the trade time, and $V_{j,t+h}$ is the terminal option value assigned to holding horizon h. The holding horizons are the close of the trade day, one, two, five, and ten trading days, and expiration. Daily gross performance is obtained by summing equation (10) across eligible trades initiated on the same trading day. The quote midpoint supplies the initial value. The measure therefore captures performance before opening execution costs.

Muravyev (2016) Quote-Based Direction follows the two-stage procedure of Muravyev (2016) used by Bryzgalova et al. (2023) in their gross-performance analysis. Trades above the NBBO midpoint are signed as purchases, and trades below it are signed as sales. Trades at the NBBO midpoint are compared with the contemporaneous PHLX local midpoint when a valid local quote is available. A trade above or below the local midpoint is signed accordingly. A trade that remains at the midpoint after the second comparison is left unsigned and excluded. The midpoint used in the successful classification step supplies the initial value $M_{j,t}$.

Terminal values are constructed from OptionMetrics closing quote midpoints. The intraday

horizon uses the closing midpoint on the trade date. The one-, two-, five-, and ten-day horizons use the first, second, fifth, and tenth subsequent daily closing observations for the same contract. When a contract has fewer remaining observations than the stated horizon, its final available closing midpoint supplies the terminal value. The expiration horizon uses the closing midpoint on the expiration date. OptionMetrics adjustment factors express terminal prices in trade-date contract units. The terminal-value history extends through August 29, 2025.

Within each holding horizon, the analysis uses a common sample across direction measures and holds fixed the trades, capped trade sizes, initial and terminal valuation prices, sample filters, and daily aggregation procedure. Exchange-Recorded Direction uses the initiator side recorded by PHLX, and GS-LASSO-Implied Direction uses out-of-sample predictions from the baseline one-month rolling specification. Decision-Tree-Implied Direction uses the predictions of the five fixed depth-three public-data trees reported in Appendix L. The trees are estimated once using January 2024 and applied without re-estimation from February through December 2024; their mean daily out-of-sample accuracy is 74.26%. The GS-LASSO direction enters this exercise as an ex-post signing measure, and its gross-performance estimates provide a downstream direction-measure comparison.

Table J1 supplements the main paper's Muravyev (2016) versus exchange-recorded comparison with the baseline GS-LASSO-implied direction and the fixed public-data decision trees.

The fixed public-data trees produce positive and statistically significant gross-performance estimates through ten trading days and an insignificant expiration estimate. Despite their relatively high classification accuracy, the trees do not recover the exchange-recorded gross-performance inference. GS-LASSO-implied gross performance is positive at every horizon; only the intraday and one-day estimates are marginally significant at the 10% level, and none is significant at the 5% level. We interpret the GS-LASSO estimates as a descriptive downstream measurement application because the baseline classifier is designed for ex-post trade signing and may use short-horizon post-trade public signals.

Table J1 Gross Performance under Exchange-Recorded, Baseline GS-LASSO, and Fixed Decision-Tree Direction

This table reports mean daily gross performance, in millions of dollars, for PHLX SLAN trades from February 1 through December 31, 2024. Positive values denote gross performance to the initiating SLAN side. Exchange-Recorded Direction uses the initiator side recorded by PHLX. GS-LASSO-Implied Direction uses out-of-sample predictions from the baseline one-month rolling specification. Decision-Tree-Implied Direction uses predictions from the five fixed depth-three public-data trees in Appendix L, whose combined mean daily out-of-sample accuracy is 74.26%. Within each holding horizon, the trades, capped trade sizes, initial and terminal valuation prices, sample filters, and aggregation procedure are held fixed across columns. The common sample contains 40,598,750 trades on 231 trading days at each horizon through ten days and 39,760,759 trades on 231 trading days at expiration. Newey–West standard errors are reported in parentheses. For each direction–horizon daily series, the lag length is selected separately using the data-dependent bandwidth rule of Newey and West (1994), without prewhitening and with the finite-sample adjustment, following Bryzgalova et al. (2023). Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Holding Horizon	Exchange-Recorded Direction	GS-LASSO-Implied Direction	Decision-Tree-Implied Direction
Intraday	−0.118	0.357*	1.659***
	(0.171)	(0.204)	(0.155)
1 Day	−0.276	0.693*	1.971***
	(0.337)	(0.411)	(0.276)
2 Days	0.090	0.894	1.906***
	(0.629)	(0.689)	(0.389)
5 Days	−0.007	2.060	2.025***
	(0.651)	(1.251)	(0.590)
10 Days	0.236	3.033	1.980**
	(0.961)	(2.004)	(0.855)
Expiration	−0.215	3.753	−0.356
	(1.193)	(2.697)	(0.927)

The common three-direction sample excludes trades left unsigned by the Muravyev (2016) procedure. For retained unsigned trades in Table J2, $M_{j,t}$ is the contemporaneous NBBO midpoint, including executions at both quote midpoints and those without an available PHLX midpoint. Table J2 removes this signability condition for exchange-recorded and GS-LASSO direction. Neither set of complete-sample estimates is statistically significant at the 10% level. We interpret the GS-LASSO estimates as a downstream measurement application because the baseline classifier is an ex-post signing procedure.

Table J2 Complete-Sample Gross Performance under Exchange-Recorded and Baseline GS-LASSO Direction

Panel A reconciles the common and complete samples. The overlap row contains trades that are both unsigned under the Muravyev (2016) procedure and missing the required terminal value. The common sample removes both exclusion sets, while the complete sample conditions only on terminal-value availability. Panel B reports mean daily gross performance, in millions of dollars, for prediction-matched PHLX SLAN trades with the required terminal value from February 1 through December 31, 2024. Each sample contains 231 trading days. Exchange-Recorded Direction and GS-LASSO-Implied Direction are defined as in Table J1. The initial valuation, capped trade sizes, terminal valuation prices, filters, aggregation procedure, and inference also follow Table J1. Newey–West standard errors are reported in parentheses. For each direction–horizon daily series, the lag length is selected separately using the data-dependent bandwidth rule of Newey and West (1994), without prewhitening and with the finite-sample adjustment, following Bryzgalova et al. (2023). Asterisks indicate statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Panel A: Observation-Count Flow

	Through Ten Days	Expiration
Prediction-matched trades	43,308,025	43,308,025
Muravyev (2016) unsigned	2,696,179	2,696,179
Terminal value unavailable	14,172	880,022
Unsigned and terminal value unavailable	1,076	28,935
Common sample	40,598,750	39,760,759
Complete sample	43,293,853	42,428,003

Panel B: Mean Daily Gross Performance

Holding Horizon	Exchange-Recorded Direction	GS-LASSO-Implied Direction
Intraday	−0.185	0.260
	(0.179)	(0.218)
1 Day	−0.325	0.630
	(0.344)	(0.434)
2 Days	0.005	0.820
	(0.686)	(0.812)
5 Days	−0.207	2.042
	(0.690)	(1.311)
10 Days	0.039	3.058
	(1.003)	(2.102)
Expiration	−0.286	3.965
	(1.251)	(2.816)

K Accuracy of the Sequential GSU Framework

We implement the sequential option-demand classification framework of Grauer et al. (2026). The GSU framework classifies customer buy and sell demand. In our matched PHLX SLAN data, accuracy is evaluated against the exchange-recorded direction of the initiating customer auction order.

We follow the combined framework described in Grauer et al. (2026, Section 4.4, pp. 15–16). The stages apply, in order, the refined displayed-size rule, the executing-exchange quote rule, the NBBO quote rule, the executing-exchange depth rule, and random assignment. Let P_i and Q_i denote transaction price and size; B_i^L and A_i^L denote the executing exchange's local best bid and offer (BBO); $Q_{B,i}^L$ and $Q_{A,i}^L$ denote displayed sizes at those local quotes; and B_i^N and A_i^N denote the national best bid and offer (NBBO). We use the numerical tolerance $\varepsilon = 10^{-5}$.

Refined displayed-size rule. In the GSU framework, the displayed-size rule is restricted to eligible standard single-leg transactions executed through the displayed limit order book.[57] Conditional on eligibility, a valid positive-width local BBO, and an execution price within that spread up to the numerical tolerance, the rule assigns

$$D_i^{(1)} = \begin{cases} +1, & |Q_i - Q_{B,i}^L| \le \varepsilon \text{ and } |Q_i - Q_{A,i}^L| > \varepsilon, \\ -1, & |Q_i - Q_{A,i}^L| \le \varepsilon \text{ and } |Q_i - Q_{B,i}^L| > \varepsilon. \end{cases}$$

Thus, a transaction is classified as buy demand when its size uniquely matches displayed bid depth and as sell demand when its size uniquely matches displayed ask depth. A transaction exits the sequence once it receives a direction.

[57] Grauer et al. (2026, Section 4.2 and footnote 4, p. 13) define standard single-leg transactions as regular and electronically executed trades, identified by OPRA LiveVol trade-condition identifiers 0 and 18. Section 4.4 (p. 15) restricts the refined displayed-size rule to these transactions. Their numerical identifiers follow LiveVol; ours follow ThetaData and consequently use a different mapping. Our sample consists of Single-Leg Auction Non-ISO (SLAN) transactions.

Executing-exchange quote rule. For each remaining transaction with a valid local quote, the second stage assigns

$$D_i^{(2)} = \begin{cases} +1, & P_i \geq (B_i^L + A_i^L)/2 + \varepsilon, \\ -1, & P_i \leq (B_i^L + A_i^L)/2 - \varepsilon. \end{cases}$$

NBBO quote rule. For each transaction remaining after the local-quote stage, the third stage applies the corresponding rule to the national quote:

$$D_i^{(3)} = \begin{cases} +1, & P_i \geq (B_i^N + A_i^N)/2 + \varepsilon, \\ -1, & P_i \leq (B_i^N + A_i^N)/2 - \varepsilon. \end{cases}$$

Executing-exchange depth rule. The fourth stage uses the displayed local depth of the remaining transactions:

$$D_i^{(4)} = \begin{cases} +1, & Q_{B,i}^L \leq Q_{A,i}^L - \varepsilon, \\ -1, & Q_{B,i}^L \geq Q_{A,i}^L + \varepsilon. \end{cases}$$

Random-assignment stage. Transactions that remain unclassified after the first four stages are randomly assigned a direction. For reported accuracy, each such transaction contributes one-half expected correct classification.

Published-eligibility implementation. Our primary specification applies the stated GSU transaction-eligibility screen at the refined displayed-size stage. Every observation in our sample is a SLAN auction transaction, so none belongs to the eligible standard-single-leg category. Consequently, no transaction enters the first-stage displayed-size rule. The displayed-size, executing-exchange quote, NBBO quote, executing-exchange depth, and random-assignment stages account for 0.00%, 87.68%, 6.08%, 5.80%, and 0.44% of transactions, respectively. Mean daily classification accuracy is 62.22%.

Expanded SLAN implementation. As a sensitivity analysis, we apply the refined displayed-size rule to every otherwise eligible SLAN transaction. The requirements of a valid positive-width local BBO, an execution within the spread, and a unique displayed-depth match remain unchanged. Under this implementation, the displayed-size, executing-exchange quote, NBBO quote, depth, and random-assignment stages account for 5.34%, 82.71%, 5.86%, 5.65%, and 0.44% of transactions, respectively. Mean daily classification accuracy is 61.80%.

Table K1 Accuracy of the GSU Framework against Exchange-Recorded Direction

This table reports summary statistics of daily classification accuracy for two implementations of the sequential framework of Grauer et al. (2026). Panel A applies the stated standard-single-leg transaction-eligibility screen at the refined displayed-size stage. No SLAN transaction enters that stage under this implementation. Panel B applies the displayed-size rule to every otherwise eligible SLAN transaction. Accuracy is evaluated against the direction of the initiating customer auction order recorded by PHLX. For each trading day and trade group, accuracy equals expected correct classifications divided by the number of transactions in the group. Transactions reaching the random-assignment stage contribute one-half expected correct classification. Results are reported for all transactions and separately for buyer-initiated transactions, seller-initiated transactions, calls, and puts. The sample contains 46,681,666 PHLX SLAN transactions over 252 trading days in 2024.

	Obs	Mean	Std	Min	5%	25%	Median	75%	95%	Max
Panel A. Stated GSU transaction-eligibility screen										
All Trades	252	62.22%	2.10%	55.46%	58.76%	60.77%	62.38%	63.73%	65.59%	67.51%
Buy Trades	252	60.06%	2.58%	53.00%	55.68%	58.22%	60.39%	61.74%	64.37%	67.15%
Sell Trades	252	64.34%	1.81%	57.34%	61.24%	63.08%	64.50%	65.52%	67.15%	68.87%
Call Trades	252	62.11%	2.00%	54.98%	58.44%	60.95%	62.35%	63.47%	65.13%	66.50%
Put Trades	252	62.34%	2.59%	54.21%	57.99%	60.65%	62.34%	64.13%	66.78%	69.39%
Panel B. Expanded displayed-size rule for SLAN transactions										
All Trades	252	61.80%	2.11%	55.29%	58.46%	60.38%	61.97%	63.33%	65.16%	67.14%
Buy Trades	252	59.67%	2.57%	52.75%	55.39%	57.76%	59.97%	61.32%	64.01%	66.76%
Sell Trades	252	63.90%	1.82%	57.23%	60.84%	62.54%	64.09%	65.13%	66.57%	68.48%
Call Trades	252	61.69%	2.01%	54.73%	58.01%	60.31%	61.90%	63.07%	64.78%	66.11%
Put Trades	252	61.93%	2.58%	54.04%	57.64%	60.18%	61.90%	63.73%	66.30%	68.95%

Under the stated GSU eligibility screen, the sequential framework achieves mean daily accuracy of 62.22%. Accuracy remains in the low 60s across buyer- and seller-initiated transactions and across calls and puts. The expanded first-stage implementation produces mean daily accuracy of 61.80%, 0.42 percentage points below the stated-screen result, with a similar pattern across the major trade groups. Both implementations therefore leave substantial classification error in the 2024 PHLX SLAN sample. We interpret these estimates as evidence on the framework's performance in this auction setting.

L Directly Implementable Public-Data Decision Trees

This section provides a small, deterministic trade-signing procedure that can be applied directly from public transaction and quote data. The procedure contains five classification and regression trees (CART), each with maximum depth three. One tree is assigned to each observable location of the option transaction price within the prevailing national best bid and offer (NBBO). The predicted class is Buy for buyer initiation and Sell for seller initiation.

L.1 Price-location groups, estimation, and evaluation

Let P_t denote the option transaction price, let B_t and A_t denote the prevailing OPRA NBBO bid and ask, let $\varepsilon = 10^{-8}$, and define

$$M_t = \frac{B_t + A_t}{2}.$$

For transactions satisfying $B_t \leq P_t \leq A_t$, the five groups are *At Bid* ($P_t = B_t$), *Below Midpoint* ($B_t < P_t < M_t - \varepsilon$), *At Midpoint* ($|P_t - M_t| \leq \varepsilon$), *Above Midpoint* ($M_t + \varepsilon < P_t < A_t$), and *At Ask* ($P_t = A_t$). The Below Midpoint and Above Midpoint groups lie strictly inside the spread. The tolerance implements numerical equality at the midpoint. This first step uses the completed OPRA transaction report and its matched contemporaneous NBBO.

This is a one-month static decision-tree specification. We estimate one tree for each group using January 2–31, 2024 and apply the five trees unchanged from February 1 through December 31, 2024. Each CART uses the Gini criterion, maximum depth three, a minimum of 1,000 training observations per terminal leaf, all available candidate variables at every split, and random seed 42. Across the eleven-month test period, the combined mean daily out-of-sample accuracy is 74.26%.

Table L1 Samples and Out-of-Sample Accuracy of the Five Fixed Decision Trees

Price-location group	Training N	OOS N	OOS share	Mean daily	Pooled
At Bid	669,456	6,191,256	14.30%	90.86%	90.97%
Below Midpoint	814,495	12,708,847	29.35%	68.90%	68.87%
At Midpoint	493,997	5,976,806	13.80%	62.16%	62.18%
Above Midpoint	750,380	12,100,644	27.94%	68.22%	68.26%
At Ask	645,313	6,330,472	14.62%	91.14%	91.24%
Combined	**3,373,641**	**43,308,025**	**100.00%**	**74.26%**	**74.21%**

Note. OOS denotes out of sample. Mean daily accuracy gives every trading day equal weight. Pooled accuracy gives every transaction equal weight.

L.2 Public information set and missing values

The candidate screen begins with 129 documented contemporaneous or pre-event variables from the 445-variable market-state file. It excludes post-trade fields, volatility-surface fields, and fields derived from the exchange-recorded initiator side. The ten variables retained by the published pruned trees are available by the transaction time and have the public OPRA constructions reported below. All timestamps are expressed in U.S. Eastern Time. The reporting-exchange BBO uses OPRA exchange-level quotes, and the NBBO uses consolidated OPRA quotes. A lagged quote is the most recent valid quote available at its indicated clock time.

The five published trees use the ten variables defined below. We use the reader-facing names and notation in Appendix M. Let

$$M^O(s) = \frac{B^O(s) + A^O(s)}{2}, \qquad S^O(s) = A^O(s) - B^O(s),$$

where the superscript O denotes the option NBBO. Superscript O,E denotes the reporting exchange option BBO. Displayed option bid and ask sizes are $D^{O,B}$ and $D^{O,A}$.

Table L2 Split Variables in the Published Pruned Trees

Symbol	Variable	Construction and timing
E	Price Position within Reporting-Exchange BBO	$E = (P_t - B_t^{O,E})/(A_t^{O,E} - B_t^{O,E})$, using the reporting exchange's prevailing option BBO.
N	Price Position within NBBO	$N = (P_t - B_t^{O})/(A_t^{O} - B_t^{O})$, using the prevailing option NBBO.
X	Spread-Scaled Execution Displacement ($t-100$ms)	$X = [P_t - M^O(t-100\text{ms})]/S^O(t-100\text{ms})$. Thus $X=0$ is the earlier midpoint and $X = \pm 0.5$ are the earlier ask and bid.
D	Displayed NBBO Depth Ratio	$D = D_t^{O,B}/D_t^{O,A}$, using the prevailing option quote.
R	Option Return ($t-1$s,$t-100$ms)	$R = \log M^O(t-100\text{ms}) - \log M^O(t-1\text{s})$.
H	Reporting-Exchange Midpoint Direction	$H = 1$ when $P_t > M_t^{O,E} + 10^{-8}$, $H = -1$ when $P_t < M_t^{O,E} - 10^{-8}$, and $H = 0$ within the tolerance. This is a public quote-position state and contains no initiator-side input.
S_{100}	Quoted Spread ($t-100$ms)	$S_{100} = S^O(t-100\text{ms})$. Tree comparisons use $\widetilde{S}_{100} = 0.01\ \text{round}(S_{100}/0.01)$, in dollars.
S_5	Quoted Spread ($t-5$s)	$S_5 = S^O(t-5\text{s})$. Tree comparisons use $\widetilde{S}_5 = 0.01\ \text{round}(S_5/0.01)$, in dollars.
W	Quoted Spread Ratio	$W = S^O(t)/M^O(t)$.
Q	Trade Size	Number of option contracts in the transaction report.

An input is missing whenever a required value is unavailable or nonfinite. Quote-based constructions require positive, uncrossed bid and ask quotes. Log returns require positive midpoints at both endpoints.

Depth ratios and imbalances require nonnegative sizes and a strictly positive denominator; all other ratios likewise require a strictly positive denominator. Reporting- exchange midpoint direction also requires an observed trade price and a valid reporting-exchange BBO.

Missing inputs remain missing during both estimation and application. At each internal node, CART learns whether a missing value follows the left or right branch. The published figures print that route in the node; "Yes" follows the displayed condition, "No" follows its complement, and each leaf prints Buy or Sell. In the direct rules, $\mathsf{Obs}(Z)$ means that Z is observed and $\mathsf{Miss}(Z)$ means that Z is missing.

L.3 Classification-equivalent direct-use rules

The initially fitted depth-3 trees use 16 distinct variables. Several internal nodes lead only to descendant leaves with the same Buy/Sell classification. Pruning every such same-class subtree preserves every hard prediction and the reported accuracy. The five published trees and their direct-use rules use ten variables:

$$\boxed{\{E, N, X, D, R, H, S_5, S_{100}, Q, W\}}.$$

Figures L1–L5 display this classification-equivalent pruned representation directly. For application, ratio, position, depth, and log-return thresholds use the decimal units in Table L2. Quoted-spread thresholds use the cent-normalized dollar spreads $\widetilde{S}_{100}$ and $\widetilde{S}_5$; trade size is in contracts. Displayed continuous thresholds use at most four decimal places.

At Bid. Predict Buy when either region A_B or region B_B holds:

$$\begin{aligned} A_B: &\quad \mathsf{Obs}(\widetilde{S}_{100}) \wedge \widetilde{S}_{100} < 0.02 \wedge [\mathsf{Miss}(Q) \vee Q < 50] \wedge [\mathsf{Miss}(X) \vee X \le 0], \\ B_B: &\quad [\mathsf{Miss}(\widetilde{S}_{100}) \vee \widetilde{S}_{100} \ge 0.02] \wedge \mathsf{Obs}(X) \wedge -0.498 < X \le -0.4045. \end{aligned}$$

All remaining observations receive the Sell prediction.

Below Midpoint. Predict Buy exactly when

$$[\mathsf{Miss}(E) \vee E > 0.4485] \wedge [\mathsf{Miss}(X) \vee X > -0.1724] \wedge [\mathsf{Miss}(D) \vee D \le 0.5234].$$

All remaining observations receive the Sell prediction.

At Midpoint. Predict Buy exactly when D is observed and

$$D \leq 0.4215 \quad \vee \quad [0.4215 < D \leq 0.9856 \wedge \mathsf{Obs}(H) \wedge H \geq 0].$$

All remaining observations, including a missing D, receive the Sell prediction.

Above Midpoint. The prediction is Sell in either of the following regions:

$$\begin{aligned} A_A: \quad & \mathsf{Obs}(E) \wedge E \leq 0.5554 \wedge [\mathsf{Miss}(N) \vee N \leq 0.648] \\ & \wedge [\mathsf{Miss}(D) \vee D > 1.4395], \\ B_A: \quad & [\mathsf{Miss}(E) \vee E > 0.5554] \wedge \mathsf{Obs}(\widetilde{S}_5) \wedge \widetilde{S}_5 \geq 0.11 \\ & \wedge \mathsf{Obs}(R) \wedge R > 0.0152. \end{aligned}$$

All remaining observations receive the Buy prediction.

At Ask. The prediction is Sell in either of the following regions:

$$\begin{aligned} A_K: \quad & \mathsf{Obs}(\widetilde{S}_{100}) \wedge \widetilde{S}_{100} < 0.02 \wedge [\mathsf{Miss}(Q) \vee Q < 50] \\ & \wedge [\mathsf{Miss}(X) \vee X > 0], \\ B_K: \quad & [\mathsf{Miss}(\widetilde{S}_{100}) \vee \widetilde{S}_{100} \geq 0.02] \wedge \mathsf{Obs}(X) \wedge X \leq 0.4911 \\ & \wedge \mathsf{Obs}(W) \wedge W > 0.2946. \end{aligned}$$

All remaining observations receive the Buy prediction.

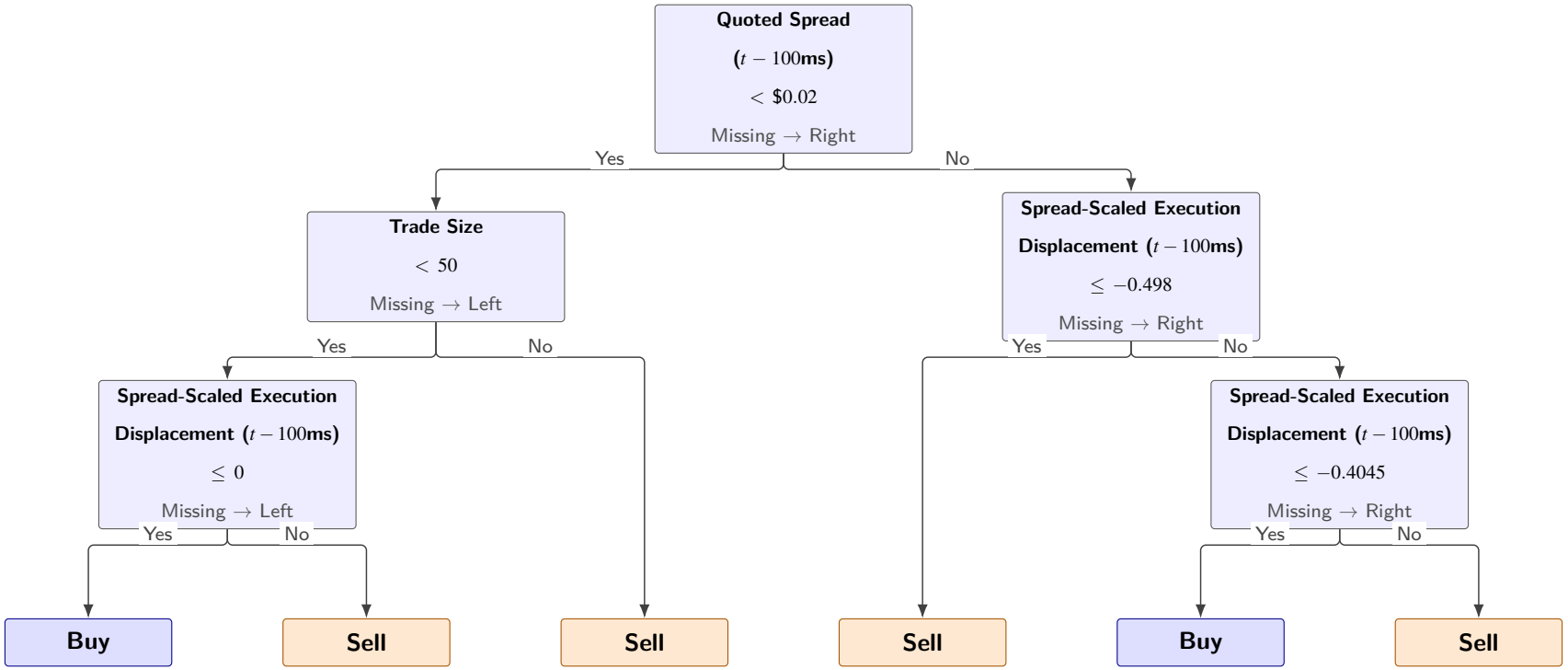


Figure L1. Pruned depth-3 public-data tree for At Bid transactions.

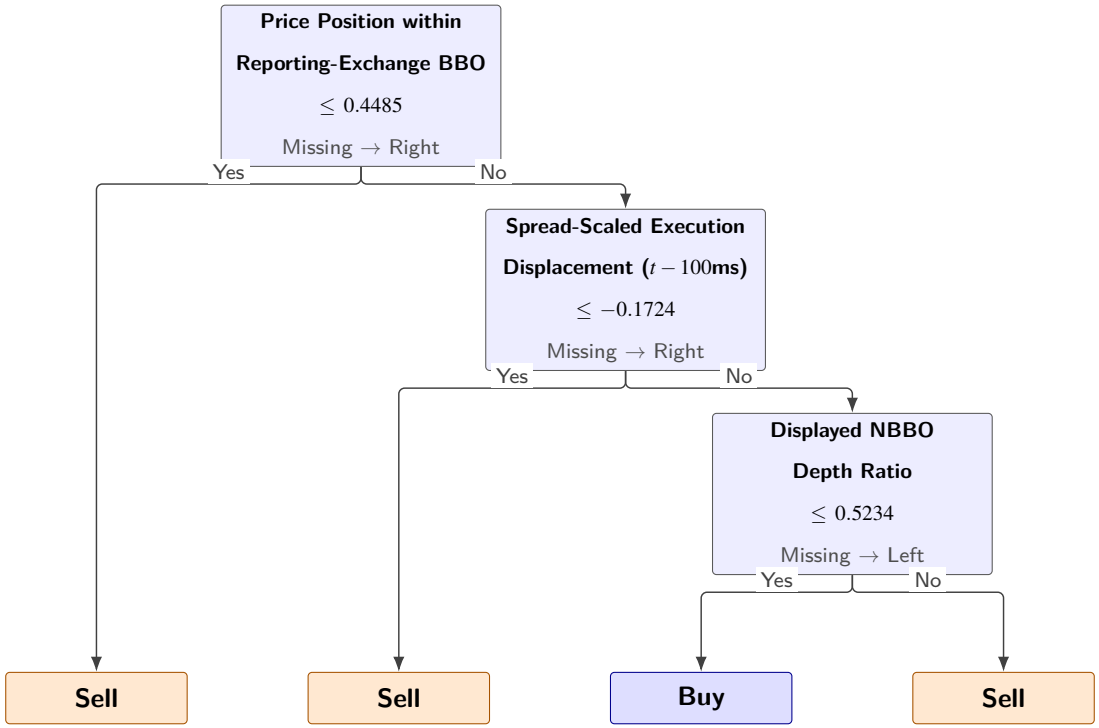


Figure L2. Pruned depth-3 public-data tree for Below Midpoint transactions.

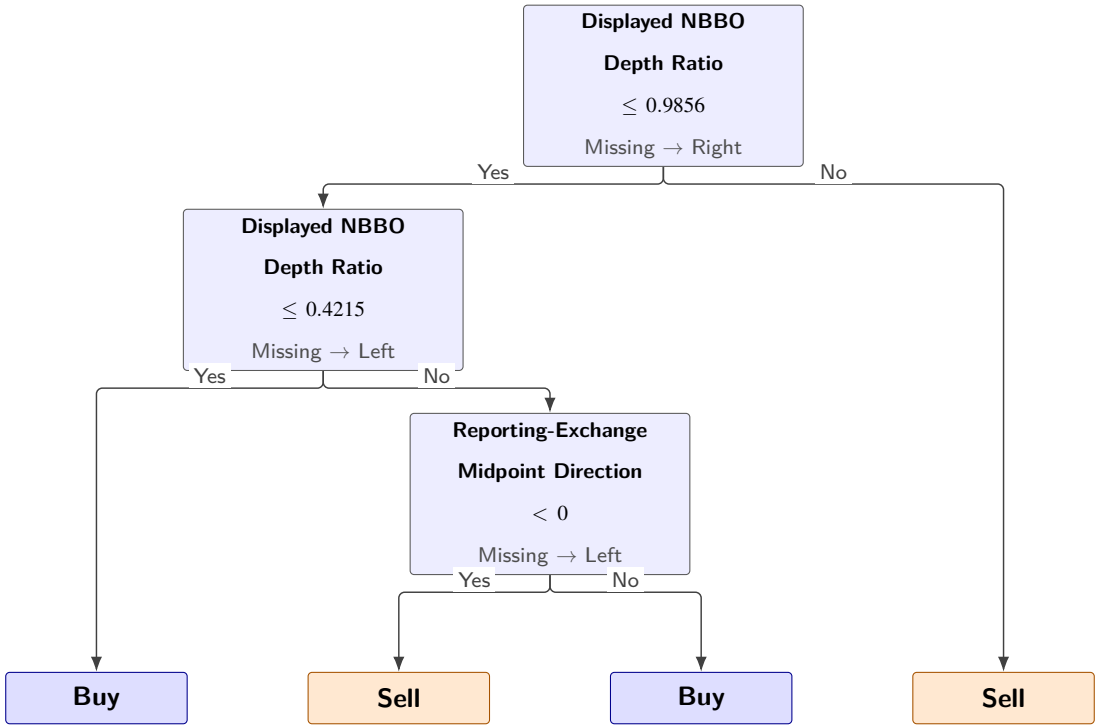


Figure L3. Pruned depth-3 public-data tree for At Midpoint transactions.

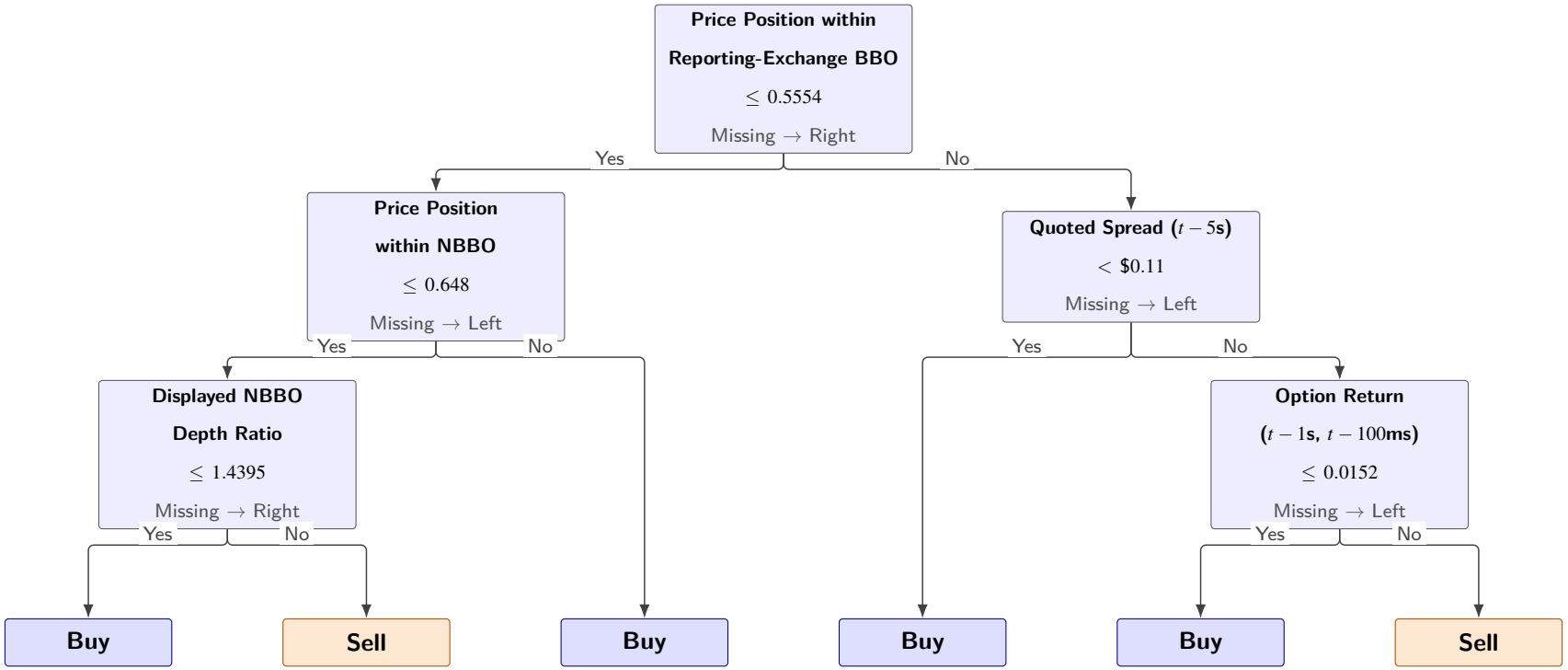


Figure L4. Pruned depth-3 public-data tree for Above Midpoint transactions.

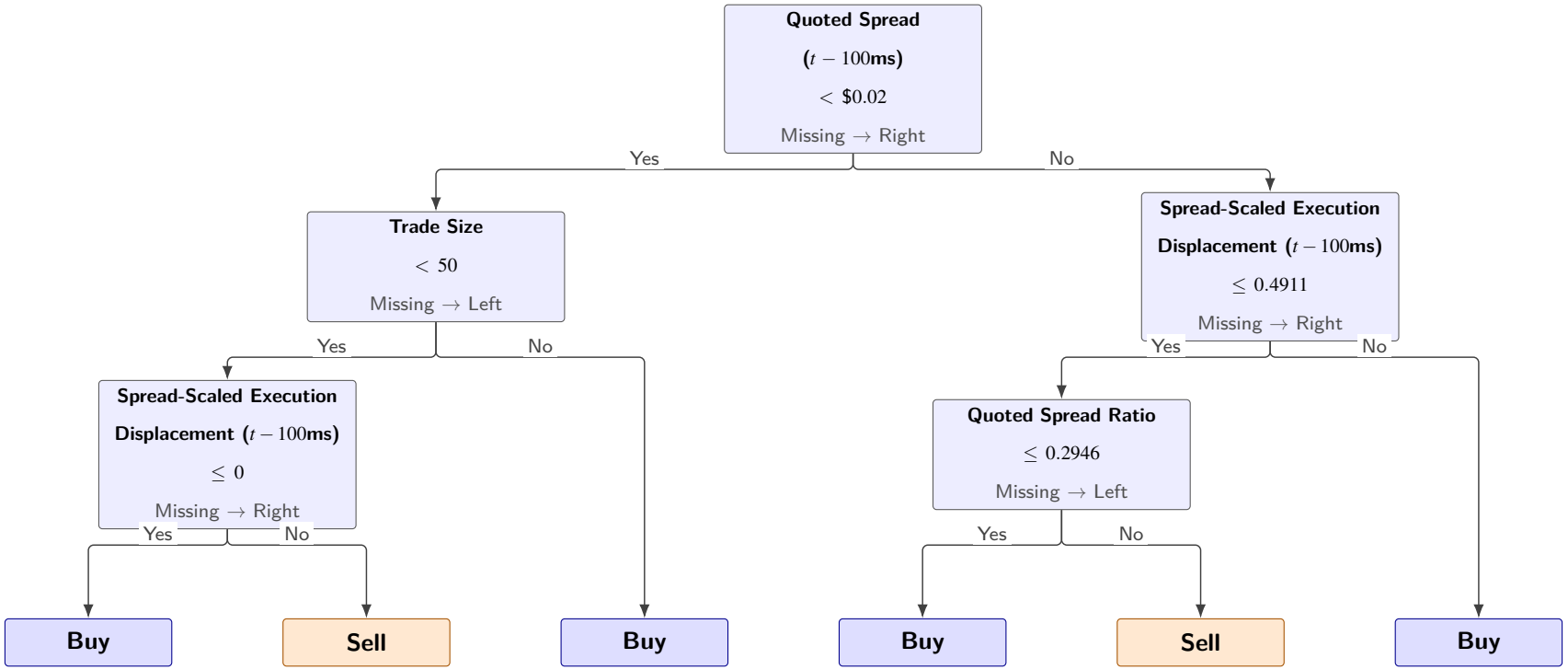


Figure L5. Pruned depth-3 public-data tree for At Ask transactions.

M Feature Definitions for the GS-LASSO Model

The baseline calendar-month GS-LASSO classifier uses 445 market-state variables organized into six groups: contract and session characteristics; option execution and current liquidity; option quote history; underlying quote and return dynamics; implied volatility, Greeks, and SpiderRock analytics; and underlying TAQ trade flow. Tables M1–M6 provide the economic name and construction of each input.

Throughout this appendix, t denotes the option-print timestamp, t^- and t^+ denote the matched pre-event and post-event quote observations, P_t is the option execution price, Q_t is trade size, K is strike price, and U_t^R is the underlying reference price matched to the option trade. The call/put orientation is $\chi = 2C - 1$, with $\chi = +1$ for calls and $\chi = -1$ for puts.

For asset $X \in \{O, U\}$, $B^X(s)$, $A^X(s)$, $M^X(s) = [B^X(s) + A^X(s)]/2$, and $S^X(s) = A^X(s) - B^X(s)$ denote the bid, ask, midpoint, and quoted spread at time s. Midpoint log returns are $r^X(a,b) = \log M^X(b) - \log M^X(a)$. Quotes at named horizons are matched to the corresponding event times. Superscripts V and E identify SpiderRock-reconstructed national quotes and reporting-exchange quotes, respectively; the same convention applies to midpoints, spreads, and displayed depth. $D^{X,B}(s)$ and $D^{X,A}(s)$ denote displayed bid and ask depth. N_t^B and N_t^A count venues at the national best bid and offer. Subscript 2 identifies the second displayed quote level.

Quote-derived measures are constructed from valid two-sided markets. The ex-post signing specification includes public option, underlying, and TAQ observations at the positive horizons listed in the tables.

M.1 Contract, Security, and Time Context

This block contains contract terms, security classification, underlying-security identifiers, and session-time controls. Representative features include Strike Price, Time to Expiration, VXX Midpoint, Large-Print Indicator, and Seconds Since Market Open.

Table M1 Contract, Security, and Time Features

Feature	Description
Contract Terms	
Strike Price	Option strike price in dollars, K.
Call/Put Indicator	Indicator C equal to one for a call and zero for a put.
Maturity and Security Classification	
Time to Expiration	ACT/365 time from the option print to 4:00 p.m. Eastern Time on the expiration date, T.
Security Type	Indicator distinguishing stock options from ETF options.
Underlying-Security Identifiers	
Underlying Symbol from Option-Series Record	Underlying symbol recorded with the option series.
Expiration Year	Calendar year of the option expiration date.
Expiration Month	Calendar month of the option expiration date.
Expiration Day	Calendar day of the option expiration date.
Underlying Symbol from Underlying-Security Record	Underlying symbol recorded with the underlying-security record.
Market and Valuation Context	
VXX Midpoint	$(B_t^{\mathrm{VXX}} + A_t^{\mathrm{VXX}})/2$, using the VXX bid and offer matched to the option print.
SpiderRock Time to Expiration	SpiderRock estimate of remaining maturity in years.
Risk-Free Rate	SpiderRock-calibrated risk-free rate matched to the option print.
Trade Classification and Session Time	
Large-Print Indicator	$\mathbf{1}(Q_t > 10)$, where Q_t is option trade size in contracts.
Half-Hour Interval	Half-hour regular-session interval containing the option print, measured in Eastern Time.
Central-Time Hour	Hour of the option print in Central Time.
Central-Time Minute of Day	Minute of day of the option print in Central Time.
Seconds Since Market Open	Elapsed seconds since the 9:30 a.m. Eastern Time market open.

M.2 Option Trade, Execution, and Current Liquidity

This block contains the option execution, matched NBBO, reporting-exchange quote, displayed depth, venue location, and print-history measures observed at the auction timestamp. Representative features include Trade Size, Displayed NBBO Depth Ratio, Price Position within NBBO, and Reporting-Exchange Depth Imbalance.

Table M2 Option Trade, Execution, and Current-Liquidity Features

Feature	Description
Option Trade and Displayed NBBO	
Trade Size	Number of option contracts in the auction print, Q_t.
Execution Price	Option premium per underlying share in the auction print, P_t.
Trade Premium	$Q_t P_t$, expressed in hundreds of dollars.
Displayed NBBO Bid Size	Displayed contract depth at the matched national best bid, $D_t^{O,B}$.
Displayed NBBO Bid Price	Matched national best bid premium per underlying share, B_t^O.
Displayed NBBO Ask Size	Displayed contract depth at the matched national best offer, $D_t^{O,A}$.
Displayed NBBO Ask Price	Matched national best offer premium per underlying share, A_t^O.
Execution State at the Matched NBBO	
Displayed NBBO Midpoint	$M_t^O = (B_t^O + A_t^O)/2$.
Moneyness	K/U_t^R, the strike-to-underlying reference-price ratio.
Midpoint-Rule Direction	$+1$ when $P_t > M_t^O$, -1 when $P_t < M_t^O$, and zero at the midpoint.
Quoted Spread (t)	$S_t^O = A_t^O - B_t^O$.
Displayed NBBO Depth Ratio	$D_t^{O,B}/D_t^{O,A}$.
Midpoint Effective Spread	$\lvert P_t - M_t^O \rvert / M_t^O$.
Trade–Quote Time Difference	Elapsed time, in seconds, between the matched option quote and the option print.
Implied-Volatility Midpoint Direction	$\mathrm{sgn}[\sigma_t^T - (\sigma_t^B + \sigma_t^A)/2]$, comparing trade implied volatility with the bid–ask implied-volatility midpoint.
NBB Venue Difference	Indicator equal to one when the NBB posting venue differs from the execution venue.
NBO Venue Difference	Indicator equal to one when the NBO posting venue differs from the execution venue.
Normalized Execution and Venue Location	
Any-Side Venue Difference	Indicator equal to one when either observed NBBO side is posted away from the execution venue.

Continued on next page

Table M2 continued

Feature	Description
Quoted Spread Ratio	S_t^O / M_t^O.
Bid Depth-to-Size Ratio	$D_t^{O,B} / Q_t$.
Ask Depth-to-Size Ratio	$D_t^{O,A} / Q_t$.
Reporting-Exchange Price Location	
Reporting-Exchange BBO Midpoint	$M_t^{O,E} = (B_t^{O,E} + A_t^{O,E})/2$, using the bid and offer displayed by the reporting exchange.
Reporting-Exchange Midpoint Direction	$+1$ when $P_t > M_t^{O,E}$, -1 when $P_t < M_t^{O,E}$, and zero at the midpoint.
SpiderRock Print and Reporting-Exchange Measures	
Within-Series Print-Cluster Sequence	Sequential position of the SpiderRock print cluster within the option series.
Print-Cluster Cumulative Size	Cumulative contract size within the SpiderRock print cluster.
Cumulative Option Volume	Total option-contract volume recorded through the print timestamp.
Cumulative Print-Cancel Volume	Contract volume printed and subsequently canceled through the print timestamp.
Bid-Side Print Count	SpiderRock count of option prints classified at the bid.
Ask-Side Print Count	SpiderRock count of option prints classified at the offer.
Cumulative Bid-Side Contract Volume	SpiderRock cumulative contract volume associated with bid-side prints.
Cumulative Ask-Side Contract Volume	SpiderRock cumulative contract volume associated with offer-side prints.
Reporting-Exchange Bid Price	Bid premium displayed by the reporting exchange at the print.
Reporting-Exchange Ask Price	Offer premium displayed by the reporting exchange at the print.
Reporting-Exchange Bid Size	Contract depth displayed at the reporting-exchange bid.
Reporting-Exchange Ask Size	Contract depth displayed at the reporting-exchange offer.
SpiderRock-Inferred Print Location	SpiderRock classification of the option print as bid, midpoint, ask, or unclassified.
SpiderRock Option NBBO and Depth	
SpiderRock Option NBBO Bid	SpiderRock-reconstructed option NBB at the print.
SpiderRock Option NBBO Ask	SpiderRock-reconstructed option NBO at the print.
SpiderRock Option NBB Depth	Cumulative displayed contract depth at the SpiderRock-reconstructed NBB.
SpiderRock Option NBO Depth	Cumulative displayed contract depth at the SpiderRock-reconstructed NBO.
Largest-Depth NBB Venue	NBBO bid-side venue posting the largest displayed contract depth.
Largest-Depth NBO Venue	NBBO offer-side venue posting the largest displayed contract depth.

Continued on next page

Table M2 continued

Feature	Description
Largest-Venue NBB Size	Displayed NBB size posted by the largest-depth bid-side venue.
Largest-Venue NBO Size	Displayed NBO size posted by the largest-depth offer-side venue.
NBB Venue Count	Number of exchanges displaying the national best bid.
NBO Venue Count	Number of exchanges displaying the national best offer.
Second-Level Option Bid	Highest displayed option bid below the NBB.
Second-Level Option Ask	Lowest displayed option offer above the NBO.
Second-Level Bid Depth	Cumulative displayed contract depth at the second-level bid.
Second-Level Ask Depth	Cumulative displayed contract depth at the second-level offer.
Trade Size Relative to Displayed Depth	
Trade Size Relative to NBBO Depth	$Q_t/(D_t^{O,V,B}+D_t^{O,V,A})$.
Trade Size Relative to Reporting-Exchange Depth	$Q_t/(D_t^{O,E,B}+D_t^{O,E,A})$.
Current Quote and Execution Location	
SpiderRock Option NBBO Midpoint	$M_t^{O,V}=(B_t^{O,V}+A_t^{O,V})/2$.
SpiderRock Option NBBO Spread	$S_t^{O,V}=A_t^{O,V}-B_t^{O,V}$.
SpiderRock Option NBBO Spread Ratio	$S_t^{O,V}/M_t^{O,V}$.
Reporting-Exchange Quoted Spread	$S_t^{O,E}=A_t^{O,E}-B_t^{O,E}$.
Reporting-Exchange Spread Ratio	$S_t^{O,E}/M_t^{O,E}$.
NBBO Midpoint Price Displacement	$P_t-M_t^{O,V}$.
Reporting-Exchange Midpoint Price Displacement	$P_t-M_t^{O,E}$.
Price Position within NBBO	$(P_t-B_t^{O,V})/(A_t^{O,V}-B_t^{O,V})$.
Price Position within Reporting-Exchange BBO	$(P_t-B_t^{O,E})/(A_t^{O,E}-B_t^{O,E})$.
At-or-Above NBBO Midpoint Indicator	$\mathbf{1}(P_t\geq M_t^{O,V})$.
At-or-Above Reporting-Exchange Midpoint Indicator	$\mathbf{1}(P_t\geq M_t^{O,E})$.
Inside-NBBO Indicator	$\mathbf{1}(B_t^{O,V}<P_t<A_t^{O,V})$.
Inside Reporting-Exchange BBO Indicator	$\mathbf{1}(B_t^{O,E}<P_t<A_t^{O,E})$.
Depth, Imbalance, and Quote-Source Comparisons	
Matched NBB Minus SpiderRock NBB	$B_t^{O}-B_t^{O,V}$.

Continued on next page

Table M2 continued

Feature	Description
Matched NBO Minus SpiderRock NBO	$A_t^O - A_t^{O,V}$.
SpiderRock NBBO Depth Imbalance	$(D_t^{O,V,B} - D_t^{O,V,A})/(D_t^{O,V,B} + D_t^{O,V,A})$.
Reporting-Exchange Depth Imbalance	$(D_t^{O,E,B} - D_t^{O,E,A})/(D_t^{O,E,B} + D_t^{O,E,A})$.
Total NBBO Venue Count	$N_t^B + N_t^A$.
NBBO Venue-Count Imbalance	$(N_t^B - N_t^A)/(N_t^B + N_t^A)$, where N_t^B and N_t^A are the numbers of exchanges at the NBB and NBO.
Second-Level Quoted Spread	$A_{2,t}^{O,V} - B_{2,t}^{O,V}$.
Second-Level Depth Imbalance	$(D_{2,t}^{O,V,B} - D_{2,t}^{O,V,A})/(D_{2,t}^{O,V,B} + D_{2,t}^{O,V,A})$.

M.3 Option Quote History and Price Dynamics

This block contains option bids, asks, midpoints, returns, and quoted spreads from the immediate pre-trade quote through 240 seconds after the print. Representative features include Quoted Spread (t-100ms), Option Return (t-1s,t), and Spread-Scaled Execution Displacement (t-100ms).

Table M3 Option Quote-History and Price-Dynamics Features

Feature	Description
Pre-Trade Option Bid and Ask	
Option Bid (t-5s)	Displayed option bid $B^O(t-5\text{s})$.
Option Ask (t-5s)	Displayed option ask $A^O(t-5\text{s})$.
Option Bid (t-1s)	Displayed option bid $B^O(t-1\text{s})$.
Option Ask (t-1s)	Displayed option ask $A^O(t-1\text{s})$.
Option Bid (t-100ms)	Displayed option bid $B^O(t-100\text{ms})$.
Option Ask (t-100ms)	Displayed option ask $A^O(t-100\text{ms})$.
Option Bid (t^-)	Displayed option bid $B^O(t^-)$.
Option Ask (t^-)	Displayed option ask $A^O(t^-)$.
Post-Trade Option Bid and Ask	
Option Bid (t^+)	Displayed option bid $B^O(t^+)$.
Option Bid (t+1s)	Displayed option bid $B^O(t+1\text{s})$.
Option Bid (t+5s)	Displayed option bid $B^O(t+5\text{s})$.
Option Bid (t+15s)	Displayed option bid $B^O(t+15\text{s})$.
Option Bid (t+30s)	Displayed option bid $B^O(t+30\text{s})$.
Option Bid (t+60s)	Displayed option bid $B^O(t+60\text{s})$.
Option Bid (t+90s)	Displayed option bid $B^O(t+90\text{s})$.
Option Bid (t+120s)	Displayed option bid $B^O(t+120\text{s})$.
Option Bid (t+180s)	Displayed option bid $B^O(t+180\text{s})$.
Option Bid (t+240s)	Displayed option bid $B^O(t+240\text{s})$.
Option Ask (t^+)	Displayed option ask $A^O(t^+)$.
Option Ask (t+1s)	Displayed option ask $A^O(t+1\text{s})$.
Option Ask (t+5s)	Displayed option ask $A^O(t+5\text{s})$.
Option Ask (t+15s)	Displayed option ask $A^O(t+15\text{s})$.

Continued on next page

Table M3 continued

Feature	Description
Option Ask (t+30s)	Displayed option ask $A^O(t+30\text{s})$.
Option Ask (t+60s)	Displayed option ask $A^O(t+60\text{s})$.
Option Ask (t+90s)	Displayed option ask $A^O(t+90\text{s})$.
Option Ask (t+120s)	Displayed option ask $A^O(t+120\text{s})$.
Option Ask (t+180s)	Displayed option ask $A^O(t+180\text{s})$.
Option Ask (t+240s)	Displayed option ask $A^O(t+240\text{s})$.
Pre-Trade Option Midpoints	
Option Midpoint (t-5s)	$M^O(t-5\text{s}) = [B^O(t-5\text{s}) + A^O(t-5\text{s})]/2$.
Option Midpoint (t-1s)	$M^O(t-1\text{s}) = [B^O(t-1\text{s}) + A^O(t-1\text{s})]/2$.
Option Midpoint (t-100ms)	$M^O(t-100\text{ms}) = [B^O(t-100\text{ms}) + A^O(t-100\text{ms})]/2$.
Option Midpoint (t^-)	$M^O(t^-) = [B^O(t^-) + A^O(t^-)]/2$.
Pre-Trade Option Returns	
Option Return (t-5s,t)	$r^O(t-5\text{s},t) = \log M^O(t) - \log M^O(t-5\text{s})$.
Option Return (t-1s,t)	$r^O(t-1\text{s},t) = \log M^O(t) - \log M^O(t-1\text{s})$.
Option Return (t-100ms,t)	$r^O(t-100\text{ms},t) = \log M^O(t) - \log M^O(t-100\text{ms})$.
Option Return (t^-,t)	$r^O(t^-,t) = \log M^O(t) - \log M^O(t^-)$.
Pre-Trade Option Interval Returns	
Option Return (t-1s,t-100ms)	$r^O(t-1\text{s},t-100\text{ms}) = \log M^O(t-100\text{ms}) - \log M^O(t-1\text{s})$.
Option Return (t-5s,t-1s)	$r^O(t-5\text{s},t-1\text{s}) = \log M^O(t-1\text{s}) - \log M^O(t-5\text{s})$.
Pre-Trade Absolute Option Returns	
Absolute Option Return (t-5s,t)	$\lvert r^O(t-5\text{s},t)\rvert$.
Absolute Option Return (t-1s,t)	$\lvert r^O(t-1\text{s},t)\rvert$.
Absolute Option Return (t-100ms,t)	$\lvert r^O(t-100\text{ms},t)\rvert$.
Absolute Option Return (t^-,t)	$\lvert r^O(t^-,t)\rvert$.
Post-Trade Option Midpoints	
Option Midpoint (t^+)	$M^O(t^+) = [B^O(t^+) + A^O(t^+)]/2$.
Option Midpoint (t+1s)	$M^O(t+1\text{s}) = [B^O(t+1\text{s}) + A^O(t+1\text{s})]/2$.
Option Midpoint (t+5s)	$M^O(t+5\text{s}) = [B^O(t+5\text{s}) + A^O(t+5\text{s})]/2$.
Option Midpoint (t+15s)	$M^O(t+15\text{s}) = [B^O(t+15\text{s}) + A^O(t+15\text{s})]/2$.
Option Midpoint (t+30s)	$M^O(t+30\text{s}) = [B^O(t+30\text{s}) + A^O(t+30\text{s})]/2$.

Continued on next page

Table M3 continued

Feature	Description
Option Midpoint (t+60s)	$M^O(t+60\text{s}) = [B^O(t+60\text{s}) + A^O(t+60\text{s})]/2.$
Option Midpoint (t+90s)	$M^O(t+90\text{s}) = [B^O(t+90\text{s}) + A^O(t+90\text{s})]/2.$
Option Midpoint (t+120s)	$M^O(t+120\text{s}) = [B^O(t+120\text{s}) + A^O(t+120\text{s})]/2.$
Option Midpoint (t+180s)	$M^O(t+180\text{s}) = [B^O(t+180\text{s}) + A^O(t+180\text{s})]/2.$
Option Midpoint (t+240s)	$M^O(t+240\text{s}) = [B^O(t+240\text{s}) + A^O(t+240\text{s})]/2.$
Post-Trade Option Returns	
Option Return (t,t$^+$)	$r^O(t,t^+) = \log M^O(t^+) - \log M^O(t).$
Option Return (t,t+1s)	$r^O(t,t+1\text{s}) = \log M^O(t+1\text{s}) - \log M^O(t).$
Option Return (t,t+5s)	$r^O(t,t+5\text{s}) = \log M^O(t+5\text{s}) - \log M^O(t).$
Option Return (t,t+15s)	$r^O(t,t+15\text{s}) = \log M^O(t+15\text{s}) - \log M^O(t).$
Option Return (t,t+30s)	$r^O(t,t+30\text{s}) = \log M^O(t+30\text{s}) - \log M^O(t).$
Option Return (t,t+60s)	$r^O(t,t+60\text{s}) = \log M^O(t+60\text{s}) - \log M^O(t).$
Option Return (t,t+90s)	$r^O(t,t+90\text{s}) = \log M^O(t+90\text{s}) - \log M^O(t).$
Option Return (t,t+120s)	$r^O(t,t+120\text{s}) = \log M^O(t+120\text{s}) - \log M^O(t).$
Option Return (t,t+180s)	$r^O(t,t+180\text{s}) = \log M^O(t+180\text{s}) - \log M^O(t).$
Option Return (t,t+240s)	$r^O(t,t+240\text{s}) = \log M^O(t+240\text{s}) - \log M^O(t).$
Post-Trade Absolute Option Returns	
Absolute Option Return (t,t$^+$)	$\lvert r^O(t,t^+)\rvert.$
Absolute Option Return (t,t+1s)	$\lvert r^O(t,t+1\text{s})\rvert.$
Absolute Option Return (t,t+5s)	$\lvert r^O(t,t+5\text{s})\rvert.$
Absolute Option Return (t,t+15s)	$\lvert r^O(t,t+15\text{s})\rvert.$
Absolute Option Return (t,t+30s)	$\lvert r^O(t,t+30\text{s})\rvert.$
Absolute Option Return (t,t+60s)	$\lvert r^O(t,t+60\text{s})\rvert.$
Absolute Option Return (t,t+90s)	$\lvert r^O(t,t+90\text{s})\rvert.$
Absolute Option Return (t,t+120s)	$\lvert r^O(t,t+120\text{s})\rvert.$
Absolute Option Return (t,t+180s)	$\lvert r^O(t,t+180\text{s})\rvert.$
Absolute Option Return (t,t+240s)	$\lvert r^O(t,t+240\text{s})\rvert.$
Pre-Trade Option Quoted Spreads	
Quoted Spread (t-100ms)	$S^O(t-100\text{ms}) = A^O(t-100\text{ms}) - B^O(t-100\text{ms}).$
Quoted Spread (t$^-$)	$S^O(t^-) = A^O(t^-) - B^O(t^-).$

Continued on next page

Table M3 continued

Feature	Description
Quoted Spread (t-1s)	$S^O(t-1\text{s}) = A^O(t-1\text{s}) - B^O(t-1\text{s})$.
Quoted Spread (t-5s)	$S^O(t-5\text{s}) = A^O(t-5\text{s}) - B^O(t-5\text{s})$.
Pre-Trade Option Spread Dynamics	
Quoted-Spread Change (t-100ms,t-)	$[S^O(t^-) - S^O(t-100\text{ms})]/S^O(t-100\text{ms})$.
Quoted-Spread Change (t-1s,t-100ms)	$[S^O(t-100\text{ms}) - S^O(t-1\text{s})]/S^O(t-1\text{s})$.
Relative Quoted Spread (t-100ms)	$S^O(t-100\text{ms})/M^O(t-100\text{ms})$.
Spread-Scaled Execution Displacement (t-100ms)	$[P_t - M^O(t-100\text{ms})]/S^O(t-100\text{ms})$.
Post-Trade Option Quoted Spreads	
Quoted Spread (t^+)	$S^O(t^+) = A^O(t^+) - B^O(t^+)$.
Quoted Spread (t+1s)	$S^O(t+1\text{s}) = A^O(t+1\text{s}) - B^O(t+1\text{s})$.
Quoted Spread (t+5s)	$S^O(t+5\text{s}) = A^O(t+5\text{s}) - B^O(t+5\text{s})$.
Quoted Spread (t+15s)	$S^O(t+15\text{s}) = A^O(t+15\text{s}) - B^O(t+15\text{s})$.
Quoted Spread (t+30s)	$S^O(t+30\text{s}) = A^O(t+30\text{s}) - B^O(t+30\text{s})$.
Quoted Spread (t+60s)	$S^O(t+60\text{s}) = A^O(t+60\text{s}) - B^O(t+60\text{s})$.
Quoted Spread (t+90s)	$S^O(t+90\text{s}) = A^O(t+90\text{s}) - B^O(t+90\text{s})$.
Quoted Spread (t+120s)	$S^O(t+120\text{s}) = A^O(t+120\text{s}) - B^O(t+120\text{s})$.
Quoted Spread (t+180s)	$S^O(t+180\text{s}) = A^O(t+180\text{s}) - B^O(t+180\text{s})$.
Quoted Spread (t+240s)	$S^O(t+240\text{s}) = A^O(t+240\text{s}) - B^O(t+240\text{s})$.

M.4 Underlying Quote and Stock-Return Dynamics

This block applies the same quote-history structure to the underlying security and adds call/put-oriented returns. It includes underlying bid and ask snapshots, midpoint returns, absolute returns, and quoted spreads from the pre-trade interval through 240 seconds after the print. Representative features include Underlying Return (t-1s,t), Call/Put-Oriented Underlying Return (t-1s,t), and Underlying Quoted Spread (t).

Table M4 Underlying Quote and Return-Dynamics Features

Feature	Description
Underlying Quote at the Option Print	
SpiderRock Underlying Bid (t)	SpiderRock-reconstructed underlying NBB at the option-print timestamp, $B^U(t) \equiv B_t^{U,V}$.
SpiderRock Underlying Ask (t)	SpiderRock-reconstructed underlying NBO at the option-print timestamp, $A^U(t) \equiv A_t^{U,V}$.
Underlying Price (t)	SpiderRock estimate of the underlying price at the option-print timestamp, $\widetilde{U}_t^V$.
Pre-Trade Underlying Bid and Ask	
Underlying Bid (t-5s)	Displayed underlying bid $B^U(t-5\text{s})$.
Underlying Ask (t-5s)	Displayed underlying ask $A^U(t-5\text{s})$.
Underlying Bid (t-1s)	Displayed underlying bid $B^U(t-1\text{s})$.
Underlying Ask (t-1s)	Displayed underlying ask $A^U(t-1\text{s})$.
Underlying Bid (t-100ms)	Displayed underlying bid $B^U(t-100\text{ms})$.
Underlying Ask (t-100ms)	Displayed underlying ask $A^U(t-100\text{ms})$.
Underlying Bid (t^-)	Displayed underlying bid $B^U(t^-)$.
Underlying Ask (t^-)	Displayed underlying ask $A^U(t^-)$.
Post-Trade Underlying Bid and Ask	
Underlying Bid (t^+)	Displayed underlying bid $B^U(t^+)$.
Underlying Bid (t+1s)	Displayed underlying bid $B^U(t+1\text{s})$.
Underlying Bid (t+5s)	Displayed underlying bid $B^U(t+5\text{s})$.
Underlying Bid (t+15s)	Displayed underlying bid $B^U(t+15\text{s})$.
Underlying Bid (t+30s)	Displayed underlying bid $B^U(t+30\text{s})$.
Underlying Bid (t+60s)	Displayed underlying bid $B^U(t+60\text{s})$.
Underlying Bid (t+90s)	Displayed underlying bid $B^U(t+90\text{s})$.
Underlying Bid (t+120s)	Displayed underlying bid $B^U(t+120\text{s})$.

Continued on next page

Table M4 continued

Feature	Description
Underlying Bid (t+180s)	Displayed underlying bid $B^U(t+180\text{s})$.
Underlying Bid (t+240s)	Displayed underlying bid $B^U(t+240\text{s})$.
Underlying Ask (t^+)	Displayed underlying ask $A^U(t^+)$.
Underlying Ask (t+1s)	Displayed underlying ask $A^U(t+1\text{s})$.
Underlying Ask (t+5s)	Displayed underlying ask $A^U(t+5\text{s})$.
Underlying Ask (t+15s)	Displayed underlying ask $A^U(t+15\text{s})$.
Underlying Ask (t+30s)	Displayed underlying ask $A^U(t+30\text{s})$.
Underlying Ask (t+60s)	Displayed underlying ask $A^U(t+60\text{s})$.
Underlying Ask (t+90s)	Displayed underlying ask $A^U(t+90\text{s})$.
Underlying Ask (t+120s)	Displayed underlying ask $A^U(t+120\text{s})$.
Underlying Ask (t+180s)	Displayed underlying ask $A^U(t+180\text{s})$.
Underlying Ask (t+240s)	Displayed underlying ask $A^U(t+240\text{s})$.
Pre-Trade Underlying Midpoints	
Underlying Midpoint (t-5s)	$M^U(t-5\text{s}) = [B^U(t-5\text{s}) + A^U(t-5\text{s})]/2$.
Underlying Midpoint (t-1s)	$M^U(t-1\text{s}) = [B^U(t-1\text{s}) + A^U(t-1\text{s})]/2$.
Underlying Midpoint (t-100ms)	$M^U(t-100\text{ms}) = [B^U(t-100\text{ms}) + A^U(t-100\text{ms})]/2$.
Underlying Midpoint (t^-)	$M^U(t^-) = [B^U(t^-) + A^U(t^-)]/2$.
Pre-Trade Underlying Returns	
Underlying Return (t-5s,t)	$r^U(t-5\text{s},t) = \log M^U(t) - \log M^U(t-5\text{s})$.
Underlying Return (t-1s,t)	$r^U(t-1\text{s},t) = \log M^U(t) - \log M^U(t-1\text{s})$.
Underlying Return (t-100ms,t)	$r^U(t-100\text{ms},t) = \log M^U(t) - \log M^U(t-100\text{ms})$.
Underlying Return (t^-,t)	$r^U(t^-,t) = \log M^U(t) - \log M^U(t^-)$.
Pre-Trade Absolute Underlying Returns	
Absolute Underlying Return (t-5s,t)	$\lvert r^U(t-5\text{s},t)\rvert$.
Absolute Underlying Return (t-1s,t)	$\lvert r^U(t-1\text{s},t)\rvert$.
Absolute Underlying Return (t-100ms,t)	$\lvert r^U(t-100\text{ms},t)\rvert$.
Absolute Underlying Return (t^-,t)	$\lvert r^U(t^-,t)\rvert$.
Post-Trade Underlying Midpoints	
Underlying Midpoint (t^+)	$M^U(t^+) = [B^U(t^+) + A^U(t^+)]/2$.

Continued on next page

Table M4 continued

Feature	Description
Underlying Midpoint (t+1s)	$M^U(t+1\text{s}) = [B^U(t+1\text{s}) + A^U(t+1\text{s})]/2.$
Underlying Midpoint (t+5s)	$M^U(t+5\text{s}) = [B^U(t+5\text{s}) + A^U(t+5\text{s})]/2.$
Underlying Midpoint (t+15s)	$M^U(t+15\text{s}) = [B^U(t+15\text{s}) + A^U(t+15\text{s})]/2.$
Underlying Midpoint (t+30s)	$M^U(t+30\text{s}) = [B^U(t+30\text{s}) + A^U(t+30\text{s})]/2.$
Underlying Midpoint (t+60s)	$M^U(t+60\text{s}) = [B^U(t+60\text{s}) + A^U(t+60\text{s})]/2.$
Underlying Midpoint (t+90s)	$M^U(t+90\text{s}) = [B^U(t+90\text{s}) + A^U(t+90\text{s})]/2.$
Underlying Midpoint (t+120s)	$M^U(t+120\text{s}) = [B^U(t+120\text{s}) + A^U(t+120\text{s})]/2.$
Underlying Midpoint (t+180s)	$M^U(t+180\text{s}) = [B^U(t+180\text{s}) + A^U(t+180\text{s})]/2.$
Underlying Midpoint (t+240s)	$M^U(t+240\text{s}) = [B^U(t+240\text{s}) + A^U(t+240\text{s})]/2.$
Post-Trade Underlying Returns	
Underlying Return (t,t$^+$)	$r^U(t,t^+) = \log M^U(t^+) - \log M^U(t).$
Underlying Return (t,t+1s)	$r^U(t,t+1\text{s}) = \log M^U(t+1\text{s}) - \log M^U(t).$
Underlying Return (t,t+5s)	$r^U(t,t+5\text{s}) = \log M^U(t+5\text{s}) - \log M^U(t).$
Underlying Return (t,t+15s)	$r^U(t,t+15\text{s}) = \log M^U(t+15\text{s}) - \log M^U(t).$
Underlying Return (t,t+30s)	$r^U(t,t+30\text{s}) = \log M^U(t+30\text{s}) - \log M^U(t).$
Underlying Return (t,t+60s)	$r^U(t,t+60\text{s}) = \log M^U(t+60\text{s}) - \log M^U(t).$
Underlying Return (t,t+90s)	$r^U(t,t+90\text{s}) = \log M^U(t+90\text{s}) - \log M^U(t).$
Underlying Return (t,t+120s)	$r^U(t,t+120\text{s}) = \log M^U(t+120\text{s}) - \log M^U(t).$
Underlying Return (t,t+180s)	$r^U(t,t+180\text{s}) = \log M^U(t+180\text{s}) - \log M^U(t).$
Underlying Return (t,t+240s)	$r^U(t,t+240\text{s}) = \log M^U(t+240\text{s}) - \log M^U(t).$
Post-Trade Absolute Underlying Returns	
Absolute Underlying Return (t,t$^+$)	$\lvert r^U(t,t^+)\rvert.$
Absolute Underlying Return (t,t+1s)	$\lvert r^U(t,t+1\text{s})\rvert.$
Absolute Underlying Return (t,t+5s)	$\lvert r^U(t,t+5\text{s})\rvert.$
Absolute Underlying Return (t,t+15s)	$\lvert r^U(t,t+15\text{s})\rvert.$
Absolute Underlying Return (t,t+30s)	$\lvert r^U(t,t+30\text{s})\rvert.$
Absolute Underlying Return (t,t+60s)	$\lvert r^U(t,t+60\text{s})\rvert.$
Absolute Underlying Return (t,t+90s)	$\lvert r^U(t,t+90\text{s})\rvert.$
Absolute Underlying Return (t,t+120s)	$\lvert r^U(t,t+120\text{s})\rvert.$
Absolute Underlying Return (t,t+180s)	$\lvert r^U(t,t+180\text{s})\rvert.$

Continued on next page

Table M4 continued

Feature	Description
Absolute Underlying Return (t,t+240s)	$\lvert r^U(t,t+240\text{s})\rvert$.
Call/Put-Oriented Underlying Returns	
Call/Put-Oriented Underlying Return (t-5s,t)	$\chi r^U(t-5\text{s},t)$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t-1s,t)	$\chi r^U(t-1\text{s},t)$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t-100ms,t)	$\chi r^U(t-100\text{ms},t)$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t^-,t)	$\chi r^U(t^-,t)$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t^+)	$\chi r^U(t,t^+)$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t+1s)	$\chi r^U(t,t+1\text{s})$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t+5s)	$\chi r^U(t,t+5\text{s})$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t+15s)	$\chi r^U(t,t+15\text{s})$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t+30s)	$\chi r^U(t,t+30\text{s})$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t+60s)	$\chi r^U(t,t+60\text{s})$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t+90s)	$\chi r^U(t,t+90\text{s})$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t+120s)	$\chi r^U(t,t+120\text{s})$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t+180s)	$\chi r^U(t,t+180\text{s})$, where $\chi = 2C-1$.
Call/Put-Oriented Underlying Return (t,t+240s)	$\chi r^U(t,t+240\text{s})$, where $\chi = 2C-1$.
Pre-Trade Underlying Quoted Spreads	
Underlying Quoted Spread (t-100ms)	$S^U(t-100\text{ms}) = A^U(t-100\text{ms}) - B^U(t-100\text{ms})$.
Underlying Quoted Spread (t^-)	$S^U(t^-) = A^U(t^-) - B^U(t^-)$.
Underlying Quoted Spread (t-1s)	$S^U(t-1\text{s}) = A^U(t-1\text{s}) - B^U(t-1\text{s})$.
Underlying Quoted Spread (t-5s)	$S^U(t-5\text{s}) = A^U(t-5\text{s}) - B^U(t-5\text{s})$.
Post-Trade Underlying Quoted Spreads	
Underlying Quoted Spread (t^+)	$S^U(t^+) = A^U(t^+) - B^U(t^+)$.
Underlying Quoted Spread (t+1s)	$S^U(t+1\text{s}) = A^U(t+1\text{s}) - B^U(t+1\text{s})$.
Underlying Quoted Spread (t+5s)	$S^U(t+5\text{s}) = A^U(t+5\text{s}) - B^U(t+5\text{s})$.

Continued on next page

Table M4 continued

Feature	Description
Underlying Quoted Spread (t+15s)	$S^U(t+15\text{s}) = A^U(t+15\text{s}) - B^U(t+15\text{s})$.
Underlying Quoted Spread (t+30s)	$S^U(t+30\text{s}) = A^U(t+30\text{s}) - B^U(t+30\text{s})$.
Underlying Quoted Spread (t+60s)	$S^U(t+60\text{s}) = A^U(t+60\text{s}) - B^U(t+60\text{s})$.
Underlying Quoted Spread (t+90s)	$S^U(t+90\text{s}) = A^U(t+90\text{s}) - B^U(t+90\text{s})$.
Underlying Quoted Spread (t+120s)	$S^U(t+120\text{s}) = A^U(t+120\text{s}) - B^U(t+120\text{s})$.
Underlying Quoted Spread (t+180s)	$S^U(t+180\text{s}) = A^U(t+180\text{s}) - B^U(t+180\text{s})$.
Underlying Quoted Spread (t+240s)	$S^U(t+240\text{s}) = A^U(t+240\text{s}) - B^U(t+240\text{s})$.
Underlying Spread at the Option Print	
SpiderRock Underlying Quoted Spread (t)	$A_t^{U,V} - B_t^{U,V}$.
SpiderRock Underlying Spread-to-Reference-Price Ratio (t)	$(A_t^{U,V} - B_t^{U,V})/\widetilde{U}_t^V$.

M.5 Implied Volatility, Greeks, Surface, and SpiderRock Analytics

This block contains quote and print implied volatility, option Greeks, surface coordinates, valuation gaps, probability analytics, and one-minute forward marks. Representative features include Absolute Option Delta Elasticity, Surface Price Gap, Delta-Hedged Option Return, and Delta-Neutral PnL (t+1m).

Table M5 Implied-Volatility, Greek, Surface, and SpiderRock-Analytics Features

Feature	Description
Quote and Trade Implied Volatility	
ThetaData Trade Implied Volatility	ThetaData implied volatility associated with the option transaction, σ_t^T.
ThetaData Bid Implied Volatility	ThetaData implied volatility associated with the option bid quote, σ_t^B.
ThetaData Ask Implied Volatility	ThetaData implied volatility associated with the option ask quote, σ_t^A.
ThetaData Midpoint Implied Volatility	$(\sigma_t^B + \sigma_t^A)/2$, using bid- and ask-quote implied volatility.
Print Implied Volatility and Greeks	
Print Implied Volatility	SpiderRock model-implied volatility associated with the option print, σ_t^P.
Print Delta	SpiderRock option delta at the option print, Δ_t.
Print Gamma	SpiderRock option gamma at the option print, Γ_t.
Print Theta	SpiderRock option theta at the option print, Θ_t.
Print Vega	SpiderRock option vega at the option print, $\mathcal{V}_t$.
Print Rho	SpiderRock option rho at the option print, ρ_t.
Maturity-Scaled Implied Volatility	$\sigma_t^P \sqrt{T}$.
Greek-Scaled Return Measures	
Absolute Option Delta Elasticity	$\lvert\Delta_t\rvert \widetilde{U}_t^V / M_t^O$.
Delta-Hedged Option Return (t-1s,t)	$r^O(t-1\text{s},t) - [\lvert\Delta_t\rvert \widetilde{U}_t^V / M_t^O]\, \chi r^U(t-1\text{s},t)$.
Surface and Probability Analytics	
Implied Continuous Dividend Rate	SpiderRock-implied continuous dividend rate at the option print, q_t.
Dividends to Expiration	SpiderRock estimate of total discrete dividends paid before option expiration.
X-Delta Surface Coordinate	SpiderRock surface coordinate expressed as a delta offset.
Surface Moneyness	SpiderRock moneyness coordinate used on the option-volatility surface.
Surface Implied Volatility	SpiderRock surface-implied volatility evaluated at the option print, σ_t^{surf}.
Surface Option Price	SpiderRock surface option value evaluated at the option print, P_t^{surf}.

Continued on next page

Table M5 continued

Feature	Description
At-the-Money Surface Volatility	SpiderRock at-the-money surface volatility at the option print, σ_t^{ATM}.
One-Minute Nonnegative Option-Markout Probability	SpiderRock probability that the one-minute option mark is at least the execution price.
Moneyness and Normalized Valuation	
Relative Moneyness	$(\widetilde{U}_t^V - K)/K$.
Log Moneyness	$\log(\widetilde{U}_t^V/K)$.
Volatility-Standardized Oriented Moneyness	$\chi \log(\widetilde{U}_t^V/K)/(\sigma_t^P\sqrt{T})$.
Intrinsic-Value Share of Midpoint	$\max[\chi(\widetilde{U}_t^V - K), 0]/M_t^O$.
Surface and Print-Price Gaps	
Surface Price Gap	$P_t - P_t^{\mathrm{surf}}$, where P_t^{surf} is the SpiderRock surface option value.
Relative Surface Price Gap	$(P_t - P_t^{\mathrm{surf}})/P_t$.
Surface-to-Print Implied-Volatility Gap	$\sigma_t^{\mathrm{surf}} - \sigma_t^P$.
ATM-to-Local Surface-Volatility Gap	$\sigma_t^{\mathrm{ATM}} - \sigma_t^{\mathrm{surf}}$.
Surface Price Gap in NBBO-Spread Units	$(P_t - P_t^{\mathrm{surf}})/S_t^{O,V}$.
Symmetric Surface-to-Print IV Gap	$2(\sigma_t^{\mathrm{surf}} - \sigma_t^P)/(\lvert\sigma_t^{\mathrm{surf}}\rvert + \lvert\sigma_t^P\rvert)$.
One-Minute Forward Marks	
Option NBBO Bid (t+1m)	SpiderRock-reconstructed option NBB one minute after the print, $B^{O,V}(t+1\mathrm{m})$.
Option NBBO Ask (t+1m)	SpiderRock-reconstructed option NBO one minute after the print, $A^{O,V}(t+1\mathrm{m})$.
Underlying NBBO Bid (t+1m)	SpiderRock-reconstructed underlying NBB one minute after the print, $B^{U,V}(t+1\mathrm{m})$.
Underlying NBBO Ask (t+1m)	SpiderRock-reconstructed underlying NBO one minute after the print, $A^{U,V}(t+1\mathrm{m})$.
Underlying Price (t+1m)	SpiderRock underlying price one minute after the print.
Surface Implied Volatility (t+1m)	SpiderRock surface-implied volatility one minute after the print, $\sigma^{\mathrm{surf}}(t+1\mathrm{m})$.
Surface Option Price (t+1m)	SpiderRock surface option value one minute after the print, $P^{\mathrm{surf}}(t+1\mathrm{m})$.
Continuous Dividend Rate (t+1m)	SpiderRock continuous dividend rate one minute after the print, $q(t+1\mathrm{m})$.
Delta-Neutral PnL (t+1m)	SpiderRock one-minute delta-neutral profit-and-loss mark for the option print, $\mathrm{PnL}^{\Delta 0}(t+1\mathrm{m})$.
Derived One-Minute Mark Features	
Option NBBO Midpoint (t+1m)	$M^{O,V}(t+1\mathrm{m}) = [B^{O,V}(t+1\mathrm{m}) + A^{O,V}(t+1\mathrm{m})]/2$.

Continued on next page

Table M5 continued

Feature	Description
Option NBBO Spread (t+1m)	$A^{O,V}(t+1\text{m}) - B^{O,V}(t+1\text{m})$.
One-Minute Option Midpoint Return	$[M^{O,V}(t+1\text{m}) - M_t^{O,V}]/M_t^{O,V}$.
Underlying NBBO Midpoint (t+1m)	$M^{U,V}(t+1\text{m}) = [B^{U,V}(t+1\text{m}) + A^{U,V}(t+1\text{m})]/2$.
One-Minute Underlying Mark Return	$[M^{U,V}(t+1\text{m}) - \widetilde{U}_t^V]/\widetilde{U}_t^V$.
One-Minute Option Markout	$M^{O,V}(t+1\text{m}) - P_t$.
One-Minute Surface-Price Change	$P^{\text{surf}}(t+1\text{m}) - P_t^{\text{surf}}$.
One-Minute Surface-Volatility Change	$\sigma^{\text{surf}}(t+1\text{m}) - \sigma_t^{\text{surf}}$.
One-Minute Dividend-Rate Change	$q(t+1\text{m}) - q_t$, using the SpiderRock continuous-dividend input.
Scaled Delta-Neutral PnL (t+1m)	$\text{PnL}^{\Delta 0}(t+1\text{m})/P_t$.

M.6 Underlying TAQ Trade Flow

This block contains adjacent underlying trades, signed trade counts, signed share volume, imbalance, directional pressure, five-minute VWAP, and changes in pre-trade flow imbalance. The windows range from one millisecond to five minutes around the option print. Representative features include Underlying Trade-Count Imbalance (Pre-1s), Underlying Share-Volume Pressure (Post-100ms), Call/Put-Oriented Underlying VWAP Gap (5m), and changes in normalized pre-trade flow imbalance.

TAQ notation. For underlying TAQ trade j, τ_j, p_j, v_j, and $d_j \in \{-1,+1\}$ denote timestamp, price, share size, and Lee–Ready sign. For window W,

$$N^{\pm}(W) = \sum_{\tau_j \in W} \mathbf{1}(d_j = \pm 1), \qquad \Delta N(W) = N^{+}(W) - N^{-}(W),$$

$$V^{\pm}(W) = \sum_{\tau_j \in W} v_j \mathbf{1}(d_j = \pm 1), \quad \Delta V(W) = V^{+}(W) - V^{-}(W).$$

Pre-event and post-event windows are $W_h^- = [t-h, t)$ and $W_h^+ = [t, t+h)$. The one-millisecond event window is $W_{1\text{ms}}^0 = [t, t+1\text{ms})$. Adjacent-trade features use the last reported underlying trade before the option print and the first reported trade at or after it. Normalized trade-count and share-volume imbalances equal zero when total flow in the corresponding window is zero.

Table M6 Underlying TAQ Trade-Flow Features

Feature	Description
Adjacent Underlying Trades	
Nearest Prior Underlying TAQ Direction	Lee–Ready sign of the most recent reported underlying trade before the option print.
Nearest Underlying TAQ Direction (At or After Print)	Lee–Ready sign of the first reported underlying trade at or after the option print.
Nearest Prior Underlying Signed Size	Lee–Ready signed share size of the most recent reported underlying trade before the option print.
Nearest Underlying Signed Size (At or After Print)	Lee–Ready signed share size of the first reported underlying trade at or after the option print.
One-Millisecond Event Flow	
Underlying Buy Trade Count (Pre-1ms)	$N^{+}(W_{1\text{ms}}^-)$.

Continued on next page

Table M6 continued

Feature	Description
Underlying Sell Trade Count (Pre-1ms)	$N^{-}(W^{-}_{1\text{ms}})$.
Underlying Trade-Count Imbalance (Pre-1ms)	$\Delta N(W^{-}_{1\text{ms}})$.
Underlying Trade-Count Pressure (Pre-1ms)	$\text{sgn}[\Delta N(W^{-}_{1\text{ms}})]$.
Underlying Buy Share Volume (Pre-1ms)	$V^{+}(W^{-}_{1\text{ms}})$.
Underlying Sell Share Volume (Pre-1ms)	$V^{-}(W^{-}_{1\text{ms}})$.
Underlying Share-Volume Imbalance (Pre-1ms)	$\Delta V(W^{-}_{1\text{ms}})$.
Underlying Share-Volume Pressure (Pre-1ms)	$\text{sgn}[\Delta V(W^{-}_{1\text{ms}})]$.
Underlying Buy Trade Count (Event 1ms)	$N^{+}(W^{0}_{1\text{ms}})$.
Underlying Sell Trade Count (Event 1ms)	$N^{-}(W^{0}_{1\text{ms}})$.
Underlying Trade-Count Imbalance (Event 1ms)	$\Delta N(W^{0}_{1\text{ms}})$.
Underlying Trade-Count Pressure (Event 1ms)	$\text{sgn}[\Delta N(W^{0}_{1\text{ms}})]$.
Underlying Buy Share Volume (Event 1ms)	$V^{+}(W^{0}_{1\text{ms}})$.
Underlying Sell Share Volume (Event 1ms)	$V^{-}(W^{0}_{1\text{ms}})$.
Underlying Share-Volume Imbalance (Event 1ms)	$\Delta V(W^{0}_{1\text{ms}})$.
Underlying Share-Volume Pressure (Event 1ms)	$\text{sgn}[\Delta V(W^{0}_{1\text{ms}})]$.
100-Millisecond Pre- and Post-Trade Flow	
Underlying Buy Trade Count (Pre-100ms)	$N^{+}(W^{-}_{100\text{ms}})$.
Underlying Sell Trade Count (Pre-100ms)	$N^{-}(W^{-}_{100\text{ms}})$.
Underlying Trade-Count Imbalance (Pre-100ms)	$\Delta N(W^{-}_{100\text{ms}})$.
Underlying Trade-Count Pressure (Pre-100ms)	$\text{sgn}[\Delta N(W^{-}_{100\text{ms}})]$.
Underlying Buy Share Volume (Pre-100ms)	$V^{+}(W^{-}_{100\text{ms}})$.
Underlying Sell Share Volume (Pre-100ms)	$V^{-}(W^{-}_{100\text{ms}})$.
Underlying Share-Volume Imbalance (Pre-100ms)	$\Delta V(W^{-}_{100\text{ms}})$.

Continued on next page

Table M6 continued

Feature	Description
Underlying Share-Volume Pressure (Pre-100ms)	$\mathrm{sgn}[\Delta V(W^{-}_{100\mathrm{ms}})]$.
Underlying Buy Trade Count (Post-100ms)	$N^{+}(W^{+}_{100\mathrm{ms}})$.
Underlying Sell Trade Count (Post-100ms)	$N^{-}(W^{+}_{100\mathrm{ms}})$.
Underlying Trade-Count Imbalance (Post-100ms)	$\Delta N(W^{+}_{100\mathrm{ms}})$.
Underlying Trade-Count Pressure (Post-100ms)	$\mathrm{sgn}[\Delta N(W^{+}_{100\mathrm{ms}})]$.
Underlying Buy Share Volume (Post-100ms)	$V^{+}(W^{+}_{100\mathrm{ms}})$.
Underlying Sell Share Volume (Post-100ms)	$V^{-}(W^{+}_{100\mathrm{ms}})$.
Underlying Share-Volume Imbalance (Post-100ms)	$\Delta V(W^{+}_{100\mathrm{ms}})$.
Underlying Share-Volume Pressure (Post-100ms)	$\mathrm{sgn}[\Delta V(W^{+}_{100\mathrm{ms}})]$.
One-Second Pre- and Post-Trade Flow	
Underlying Buy Trade Count (Pre-1s)	$N^{+}(W^{-}_{1\mathrm{s}})$.
Underlying Sell Trade Count (Pre-1s)	$N^{-}(W^{-}_{1\mathrm{s}})$.
Underlying Trade-Count Imbalance (Pre-1s)	$\Delta N(W^{-}_{1\mathrm{s}})$.
Underlying Trade-Count Pressure (Pre-1s)	$\mathrm{sgn}[\Delta N(W^{-}_{1\mathrm{s}})]$.
Underlying Buy Share Volume (Pre-1s)	$V^{+}(W^{-}_{1\mathrm{s}})$.
Underlying Sell Share Volume (Pre-1s)	$V^{-}(W^{-}_{1\mathrm{s}})$.
Underlying Share-Volume Imbalance (Pre-1s)	$\Delta V(W^{-}_{1\mathrm{s}})$.
Underlying Share-Volume Pressure (Pre-1s)	$\mathrm{sgn}[\Delta V(W^{-}_{1\mathrm{s}})]$.
Underlying Buy Trade Count (Post-1s)	$N^{+}(W^{+}_{1\mathrm{s}})$.
Underlying Sell Trade Count (Post-1s)	$N^{-}(W^{+}_{1\mathrm{s}})$.
Underlying Trade-Count Imbalance (Post-1s)	$\Delta N(W^{+}_{1\mathrm{s}})$.
Underlying Trade-Count Pressure (Post-1s)	$\mathrm{sgn}[\Delta N(W^{+}_{1\mathrm{s}})]$.
Underlying Buy Share Volume (Post-1s)	$V^{+}(W^{+}_{1\mathrm{s}})$.
Underlying Sell Share Volume (Post-1s)	$V^{-}(W^{+}_{1\mathrm{s}})$.

Continued on next page

Table M6 continued

Feature	Description
Underlying Share-Volume Imbalance (Post-1s)	$\Delta V(W_{1s}^{+})$.
Underlying Share-Volume Pressure (Post-1s)	$\mathrm{sgn}[\Delta V(W_{1s}^{+})]$.
Five-Second Pre- and Post-Trade Flow	
Underlying Buy Trade Count (Pre-5s)	$N^{+}(W_{5s}^{-})$.
Underlying Sell Trade Count (Pre-5s)	$N^{-}(W_{5s}^{-})$.
Underlying Trade-Count Imbalance (Pre-5s)	$\Delta N(W_{5s}^{-})$.
Underlying Trade-Count Pressure (Pre-5s)	$\mathrm{sgn}[\Delta N(W_{5s}^{-})]$.
Underlying Buy Share Volume (Pre-5s)	$V^{+}(W_{5s}^{-})$.
Underlying Sell Share Volume (Pre-5s)	$V^{-}(W_{5s}^{-})$.
Underlying Share-Volume Imbalance (Pre-5s)	$\Delta V(W_{5s}^{-})$.
Underlying Share-Volume Pressure (Pre-5s)	$\mathrm{sgn}[\Delta V(W_{5s}^{-})]$.
Underlying Buy Trade Count (Post-5s)	$N^{+}(W_{5s}^{+})$.
Underlying Sell Trade Count (Post-5s)	$N^{-}(W_{5s}^{+})$.
Underlying Trade-Count Imbalance (Post-5s)	$\Delta N(W_{5s}^{+})$.
Underlying Trade-Count Pressure (Post-5s)	$\mathrm{sgn}[\Delta N(W_{5s}^{+})]$.
Underlying Buy Share Volume (Post-5s)	$V^{+}(W_{5s}^{+})$.
Underlying Sell Share Volume (Post-5s)	$V^{-}(W_{5s}^{+})$.
Underlying Share-Volume Imbalance (Post-5s)	$\Delta V(W_{5s}^{+})$.
Underlying Share-Volume Pressure (Post-5s)	$\mathrm{sgn}[\Delta V(W_{5s}^{+})]$.
Thirty-Second Pre- and Post-Trade Flow	
Underlying Buy Trade Count (Pre-30s)	$N^{+}(W_{30s}^{-})$.
Underlying Sell Trade Count (Pre-30s)	$N^{-}(W_{30s}^{-})$.
Underlying Trade-Count Imbalance (Pre-30s)	$\Delta N(W_{30s}^{-})$.
Underlying Trade-Count Pressure (Pre-30s)	$\mathrm{sgn}[\Delta N(W_{30s}^{-})]$.

Continued on next page

Table M6 continued

Feature	Description
Underlying Buy Share Volume (Pre-30s)	$V^{+}(W_{30\text{s}}^{-})$.
Underlying Sell Share Volume (Pre-30s)	$V^{-}(W_{30\text{s}}^{-})$.
Underlying Share-Volume Imbalance (Pre-30s)	$\Delta V(W_{30\text{s}}^{-})$.
Underlying Share-Volume Pressure (Pre-30s)	$\operatorname{sgn}[\Delta V(W_{30\text{s}}^{-})]$.
Underlying Buy Trade Count (Post-30s)	$N^{+}(W_{30\text{s}}^{+})$.
Underlying Sell Trade Count (Post-30s)	$N^{-}(W_{30\text{s}}^{+})$.
Underlying Trade-Count Imbalance (Post-30s)	$\Delta N(W_{30\text{s}}^{+})$.
Underlying Trade-Count Pressure (Post-30s)	$\operatorname{sgn}[\Delta N(W_{30\text{s}}^{+})]$.
Underlying Buy Share Volume (Post-30s)	$V^{+}(W_{30\text{s}}^{+})$.
Underlying Sell Share Volume (Post-30s)	$V^{-}(W_{30\text{s}}^{+})$.
Underlying Share-Volume Imbalance (Post-30s)	$\Delta V(W_{30\text{s}}^{+})$.
Underlying Share-Volume Pressure (Post-30s)	$\operatorname{sgn}[\Delta V(W_{30\text{s}}^{+})]$.
Sixty-Second Pre- and Post-Trade Flow	
Underlying Buy Trade Count (Pre-60s)	$N^{+}(W_{60\text{s}}^{-})$.
Underlying Sell Trade Count (Pre-60s)	$N^{-}(W_{60\text{s}}^{-})$.
Underlying Trade-Count Imbalance (Pre-60s)	$\Delta N(W_{60\text{s}}^{-})$.
Underlying Trade-Count Pressure (Pre-60s)	$\operatorname{sgn}[\Delta N(W_{60\text{s}}^{-})]$.
Underlying Buy Share Volume (Pre-60s)	$V^{+}(W_{60\text{s}}^{-})$.
Underlying Sell Share Volume (Pre-60s)	$V^{-}(W_{60\text{s}}^{-})$.
Underlying Share-Volume Imbalance (Pre-60s)	$\Delta V(W_{60\text{s}}^{-})$.
Underlying Share-Volume Pressure (Pre-60s)	$\operatorname{sgn}[\Delta V(W_{60\text{s}}^{-})]$.
Underlying Buy Trade Count (Post-60s)	$N^{+}(W_{60\text{s}}^{+})$.
Underlying Sell Trade Count (Post-60s)	$N^{-}(W_{60\text{s}}^{+})$.

Continued on next page

Table M6 continued

Feature	Description
Underlying Trade-Count Imbalance (Post-60s)	$\Delta N(W^{+}_{60s})$.
Underlying Trade-Count Pressure (Post-60s)	$\mathrm{sgn}[\Delta N(W^{+}_{60s})]$.
Underlying Buy Share Volume (Post-60s)	$V^{+}(W^{+}_{60s})$.
Underlying Sell Share Volume (Post-60s)	$V^{-}(W^{+}_{60s})$.
Underlying Share-Volume Imbalance (Post-60s)	$\Delta V(W^{+}_{60s})$.
Underlying Share-Volume Pressure (Post-60s)	$\mathrm{sgn}[\Delta V(W^{+}_{60s})]$.
Five-Minute Pre-Trade VWAP	
Underlying VWAP (Pre-5m)	$\mathrm{VWAP}(W^{-}_{5m}) = \sum_{\tau_j \in W^{-}_{5m}} p_j v_j / \sum_{\tau_j \in W^{-}_{5m}} v_j$.
Call/Put-Oriented Underlying VWAP Gap (5m)	$\chi[\mathrm{VWAP}(W^{-}_{5m}) - \widetilde{U}^{V}_{t}]$.
Five-Minute Pre-Trade Flow	
Underlying Buy Trade Count (Pre-5m)	$N^{+}(W^{-}_{5m})$.
Underlying Sell Trade Count (Pre-5m)	$N^{-}(W^{-}_{5m})$.
Underlying Trade-Count Imbalance (Pre-5m)	$\Delta N(W^{-}_{5m})$.
Underlying Trade-Count Pressure (Pre-5m)	$\mathrm{sgn}[\Delta N(W^{-}_{5m})]$.
Underlying Buy Share Volume (Pre-5m)	$V^{+}(W^{-}_{5m})$.
Underlying Sell Share Volume (Pre-5m)	$V^{-}(W^{-}_{5m})$.
Underlying Share-Volume Imbalance (Pre-5m)	$\Delta V(W^{-}_{5m})$.
Underlying Share-Volume Pressure (Pre-5m)	$\mathrm{sgn}[\Delta V(W^{-}_{5m})]$.
Changes in Pre-Trade Flow Imbalance	
Change in Normalized Trade-Count Imbalance (60s vs. 5m)	$I_N(W^{-}_{60s}) - I_N(W^{-}_{5m})$, where $I_N(W) = \Delta N(W)/[N^{+}(W) + N^{-}(W)]$.
Change in Normalized Share-Volume Imbalance (60s vs. 5m)	$I_V(W^{-}_{60s}) - I_V(W^{-}_{5m})$, where $I_V(W) = \Delta V(W)/[V^{+}(W) + V^{-}(W)]$.