Option Trade Classification*

Caroline Grauer† Philipp Schuster‡ Marliese Uhrig-Homburg§

This version: May 10, 2023

Abstract

We evaluate the performance of common stock trade classification algorithms to infer the trade direction of option trades. Using a large sample of matched intraday transactions and Open/Close data, we show that the algorithms' success to classify option trades is considerably lower than for stocks. The weak performance is due to sophisticated customers who often use limit orders instead of market orders to implement their trading strategies. These traders' behavior varies over time and across exchanges with different pricing models. We introduce new rules that enhance existing algorithms and improve classification accuracy by 9% to 47%. Applying our new rules to construct a long-short trading strategy for stocks based on option order imbalance increases Sharpe ratios from 2.65 to 4.07.

JEL classification: C10, C52, D10, G10, G20

Keywords: buyer/seller initiated trades, intraday transaction analysis, Lee and Ready algorithm, limit order, tick test, trade classification, trade direction, quote rule

*This work was supported by the Deutsche Forschungsgemeinschaft [UH 107/9-1 and SCHU 3049/5-1] as part of the research unit FOR 5230 (Financial Markets and Frictions). We thank seminar participants at the annual seminar of the research unit, University of Stuttgart, and Karlsruhe Institute of Technology, as well as Monika Gehde-Trapp for helpful comments and suggestions.

† Karlsruhe Institute of Technology, Institute for Finance, P.O. Box 6980, D-76049 Karlsruhe.
Email: caroline.grauer@kit.edu.

‡ University of Stuttgart, Department of Finance, Keplerstr. 17, D-70174 Stuttgart.
Email: philipp.schuster@bwi.uni-stuttgart.de.

§ Karlsruhe Institute of Technology, Institute for Finance, P.O. Box 6980, D-76049 Karlsruhe.
Email: uhrig@kit.edu.

1 Introduction

For a wide range of research questions in option markets, the side of the customer in a trade is of primary importance. Particularly, the trade direction is required to determine the information content of trades, the price impact of customer transactions, as well as the order imbalance and inventory accumulation of intermediaries. Important examples are studies on option demand (Gârleanu, Pedersen, and Poteshman (2009); Muravyev and Ni (2020)), option order flow (Muravyev (2016)), and option price pressures (Goyenko and Zhang (2021)). Because most option datasets do not contain information on the side of a trade, empirical studies often rely on heuristics tested in stock markets to infer trade direction from prices and quotes. The most common of these classification rules are the quote rule, the tick test, the Lee and Ready (LR, 1991) algorithm, and the Ellis, Michaely, and O'Hara (EMO, 2000) rule. Notwithstanding their wide application, little is known on how stock trade classification rules perform in option markets.

In general, existing trade classification rules are based on the understanding that market makers provide liquidity by constantly quoting bid and ask prices, and that customers trade market orders at those prices. Therefore, quote rules classify trades executed at the prevailing ask price as customer buy orders and trades executed at the prevailing bid price as customer sell orders. Several studies confirm that this simple heuristic performs quite well in the stock markets. For example, Lee and Radhakrishna (2000) show that it correctly classifies 98% of trades transacted at the bid or ask price. In contrast, for this simple heuristic, we find a much lower success rate to correctly classify customer trades in the option market that is just 26% at Nasdaq GEMX, 54% at the Chicago Board Options Exchange (CBOE), and 63% at the International Securities Exchange (ISE). Our explanation is that sophisticated customers in the option market often use limit orders instead of market orders to implement their trading strategies. If such customer limit orders cannot be executed immediately, the limit prices indicate the customers' willingness to buy at the specified bid or sell at the specified ask price. Market makers eventually fill these limit orders and trigger the trade execution. In this case, buys at the bid and sells at the ask price are systematically misclassified by the quote rule.

The prevalence of such trading strategies depends on the incentives for customers to post limit orders and is particularly strong when liquidity providing participants receive a rebate from the exchange (see, e.g., Battalio, Shkilko, and Van Ness (2016) for a related discussion on the impact of fee structures and order flow inducements). Thus, misclassifications are closely linked to the pricing model and fee schedule the exchange is using. For example, Nasdaq GEMX tries to attract customers to post limit orders with an aggressive maker-taker model,[1] explaining the extremely poor performance of the existing trade classification methods for this exchange. Moreover, there are at least two other features of option markets that make the application of stock classification rules questionable. First, options are much more illiquid than stocks with many series not recording a trade for days or weeks. For that reason, tick rules that classify trades as buys (sells) if their price is higher (lower) than the preceding or succeeding trade price are problematic. Second, option trading is spread out on 16 individual exchanges linked by a national market system. Therefore, intraday transaction data not only includes the quotes from the exchange where the trade is recorded, but also the national best bid offer (NBBO). Conceptually, it is unclear on which of those two the prevailing quote rules should be applied. Against this backdrop, it is surprising that there is just one study comparing trade classification rules in option markets, which is conducted on a small and more than twenty-five-year-old dataset (Savickas and Wilson (2003)).

We fill this gap in the literature and evaluate the success of the existing trade classification algorithms to correctly infer the direction of option trades for a large sample of intraday transactions. Compared to the one existing study, our sample is much larger, with more than 125 million option trades from two of the most important option exchanges, CBOE and ISE, and one smaller exchange, Nasdaq GEMX, and covers a more than sixteen-year period starting in May 2005.[2] In addition to the comparison of established classification rules developed on the stock market, we also design two additional rules that account for the more active role of sophisticated customers in the trading process. These rules strongly improve the performance of all existing classification algorithms. To ensure that our classification rules are not optimized and tested on the same dataset,

[1] In a maker-taker pricing model, participants that provide liquidity by posting limit orders receive a rebate, whereas traders that consume liquidity by submitting market orders are charged a fee. Gary Katz, former president and chief executive officer of the ISE, to which GEMX belonged when it was launched, highlights the attractiveness of GEMX for sophisticated customers in an interview with Traders magazine: `https://www.tradersmagazine.com/departments/options/options-report-ises-gemini-to-blend-maker-taker-with-pro-rata/`.

[2] ISE and CBOE are two of the largest option exchanges, accounting together with GEMX for about 39% of all option trades and 48% of the trading volume during our sample period.

we follow a two-step research plan. In the first step, we test existing methods and develop our improvements based just on trades recorded at the ISE between 2005 and 2017. The results of this step were registered with the Open Science Framework (OSF). In the second step, we test our algorithms out-of-sample on data from the CBOE and GEMX that have been purchased after the results of the first step were registered.[3]

Comparing the performance of trade classification algorithms requires information on the true side of the trade. To arrive at such a benchmark, we combine information from intraday trade data provided by LiveVol and daily Open/Close data from the ISE, the CBOE, or the GEMX, respectively. We take advantage of the fact that if there were only customer buy or only customer sell orders on a specific day for a given option series at one of the exchanges (ISE, CBOE, or GEMX), Open/Close data allows to directly observe the direction of all trades on that day for that specific exchange. Our three matched samples therefore consist of all intraday trades at the respective exchange under consideration on days when the trading volume of an option series is equal to either the total customer buy volume or the total customer sell volume. For these trades, market makers take the opposite position in the option trade and we assume, consistent with the previous literature, that in such cases the customer is the party with option demand. This allows us to unambiguously identify the trade direction.

Evaluating the performance of different classification algorithms, we find that the accuracy of existing methods to sign option trades as customer buys or sells from price and quote data strongly differs between the methods but also depending on the trading venue on which they are applied. We find that quote rules strongly outperform tick rules, whereas the decision on whether to apply quote rules first on the NBBO or the quotes on the exchange or vice versa has only minimal impact on the results. For the tick rule, a higher success rate can be achieved using prices across all exchanges and information from subsequent trades. These results also hold for the methods that combine quote and tick rules, of which the LR outperforms the EMO rule. On the ISE and the CBOE, the

[3] The Open Science Framework (`osf.io`) is a platform that facilitates the registration of research to increase its credibility. We created a time stamped, read-only document that contains the results from the first step, i.e., the specifications of our algorithms and the results for the ISE, and our plan for the second step. The document was uploaded on November 20, 2021. At that time, we did not have access to the CBOE and GEMX Open/Close datasets. The document is available via the following link: `https://osf.io/kj86r/?view_only=388a89b23254425a8271402e2b11fc4e`.

highest success rates can be achieved by applying quote rules first and classifying the remaining trades with a variant of the tick test, which corresponds to the LR algorithm. This way, between 60% and 64% of trades are classified correctly, which is slightly higher than using quote rules alone. Most strikingly, the performance of these standard methods for the GEMX, which employs strong incentives for customers to post limit orders, is with 31% to 34% lower than the 50% that one would expect with a purely random assignment. Overall, the accuracy of existing classification methods is considerably lower for option trades than for stocks, which is mostly between 70% and 90% (see, e.g., Lee and Ready (1991); Ellis, Michaely, and O'Hara (2000); Chakrabarty, Li, Nguyen, and Van Ness (2007)).

We hypothesize that the overall weak performance is driven by sophisticated customers often using limit orders instead of market orders when implementing their trading strategies. A large literature on optimal order submission analyzes the trade-offs between the submission of market orders and limit orders (see, e.g., Foucault, Kadan, and Kandel (2005); Goettler, Parlour, and Rajan (2005)). Results from this literature show that depending on the situation, e.g., if the trader is relatively patient and the bid-ask spread is wide, it is optimal to submit a limit order even if there is a risk that the order is never executed. In the option market with its relatively large bid-ask spreads and often infrequent updating of option bid and ask quotes, it seems plausible that such situations occur more frequently than in highly liquid equity markets (see Muravyev and Pearson (2020) for a related discussion on liquidity timing). Our analysis shows that, although sophisticated customers provide liquidity to other market participants through their limit orders, such orders are in most situations executed by a market maker and, comparing them to end-of-day prices, still leads to transaction costs paid by the customer. These costs are, however, much lower compared to the transaction costs paid by customers submitting market orders.[4] Customers considering both limit and market orders for their optimal execution strategies are not consistent with traditional classification rules that assume that customers always trade via market orders and market makers never cross the spread. Consequently, when a customer's limit order is filled by a market maker, the order is systematically misclassified by quote rules.

We identify such limit orders by comparing trade sizes with bid and ask quote sizes and propose

[4] An analysis of the profitability of customers posting market orders and limit orders is provided in Appendix A.1.

a correction to their classification. The main idea of our new "trade size rule" is that when the trade size matches exactly either the bid or ask quote size, it is likely that the quote came from a customer, the market maker found it attractive and, therefore, decided to fill it completely. We find that between 78% and 92% of all trades, for which either the bid or ask quote size corresponds to the trade size, are correctly classified using our newly proposed rule. Considering all trades in the matched dataset, our trade size rule improves the performance of the classification algorithms by about 9% for the CBOE, 11% for the ISE, and 38% to 46% at the GEMX.

Our second new rule addresses the fact that midspread trades are particularly difficult to classify, which leads to the only minor improvements of the LR algorithms compared to the quote rule. We propose an alternative to the tick test to classify midspread trades based on the comparison of bid and ask quoted depths. Using this "depth rule" leads to a further improvement of between 0.1% and 1.2% to correctly classify trades compared to the LR algorithm. Across all trades in the matched datasets, applying the trade size and the depth rule together leads to a success rate between 64% and 81%. Using our improvements, the classification precision now is much closer to the precision documented for stock trades.

Following Hu (2014), we evaluate the predictability of future stock returns based on option order imbalances, and calculate long-short portfolio returns from buying the quintile of stocks with the most positive option order imbalance and selling the portfolio with the most negative option order imbalance. Comparing the results based on order imbalance calculated with existing trade classification rules and our new classification rules, we find that the higher accuracy of our new rules translates to a higher predictability of future stock returns. Over our more than 17 year long sample period from January 2004 to June 2021, annual excess returns increase from 10.6% when existing rules are used to 13.4% for our new rules. The effect is even stronger for Sharpe ratios that increase from 2.65 to 4.07.

Our paper contributes to at least two strands of literature. First, it is closely related to the literature that evaluates the accuracy of trade classification rules for various asset classes. There exists a vast literature that tests and compares the success of competing trade classification rules to identify buy and sell orders on various markets and at different time periods. Table A1 in the Appendix provides an overview of the most important studies. The general finding from this literature is that the overall success of classification rules is relatively high, but varies widely across security markets and time periods. To the best of our knowledge, Savickas and Wilson (2003) provide the only study that examines the trade classification accuracy for option trades. Their study is based on a proprietary dataset from the CBOE covering options on 826 individual underlying assets (stocks and a few market indexes) over the period from July to December 1995. They document a much lower classification precision with options data than with stock data. The reported success rates range from 59% for the tick test to 83% for the quote rule. Our study complements the results of Savickas and Wilson (2003) by documenting a considerably lower accuracy of existing classification rules for option trades than for stock trades, using a much larger sample of option data. We further provide an economic explanation that the overall weak performance is due to the fact that in the relatively illiquid option market, sophisticated customers often use limit orders instead of market orders to implement their trading strategies. Based on this mechanism and the poor performance of the tick test, we propose two simple rules that can be used in combination with existing classification algorithms. Both rules lead to a substantial improvement in classification precision.

Second, our paper contributes to the literature that examines the impact of informed trading in the option market on price discovery in the underlying stock market. While Chan, Chung, and Fong (2002) find no evidence that option order flow has a pricing effect, Easley, O'Hara, and Srinivas (1998) and Hu (2014) report that signed trading volume predicts the underlying stock returns. Drilling down to the drivers of the influence, Cao, Chen, and Griffin (2005) provide evidence that option order imbalance becomes informative for future stock prices right before takeover announcements and Goyenko, Ornthanalai, and Tang (2015) show that option order flow predicts underlying returns primarily when option illiquidity simultaneously increases. Our results show that the improvement in classification precision translates to a higher predictability of future stock returns. A potential explanation for the mixed evidence in previous papers are the differences in the classification success rates across sample periods, option characteristics, and option exchanges that we document.

The rest of this paper is structured as follows. In Section 2, we describe the option data and sample selection process when matching LiveVol and Open/Close data and the methodology to infer the true side of a trade. Section 3 compares the performance of the common stock trade classification algorithms when applied to option trades on the ISE and introduces two new rules that strongly improve the classification success of existing methods. In Section 4, we test our new rules out-of-sample on the CBOE and the GEMX data. Section 5 provides sample splits along various dimensions and a time-series analysis of the improvements from our new rules. Section 6 applies our new trade classification algorithms to test the predictability of future stock returns based on option order imbalances. Finally, Section 7 concludes the paper.

2 Data and Benchmark Classification Methodology

2.1 Data

Our empirical analyses use option data from three different sources. We obtain intraday option price and quote data from LiveVol, end-of-day buy and sell trading volumes categorized by trader type from Open/Close Trade Profiles of three individual option exchanges, and option and underlying characteristics from Ivy DB OptionMetrics.

LiveVol provides intraday transaction-level option data for all option trades on all U.S. exchanges. This includes the execution price, the trading volume, the national best bid and offer (NBBO) quotes at the time of the trade, the individual quote and quote sizes for each exchange on which the option is quoted, and information on the exchange on which the trade is executed. We apply a minimal list of filters to this data. We filter out option trades with a trading price less than or equal to zero. We also remove trades with negative or zero volume and those whose trading volume exceeds 10 million contracts. Furthermore, we delete entries with multiple underlying symbols for the same root and other duplicates along with any cancelled trades.

Our three Open/Close datasets from the ISE, the CBOE, and the GEMX each contain daily trading volumes for the option series traded at the respective exchange. The volume is broken down into buys and sells, whether the trades open new or close existing option positions and is categorized by customer, professional customer, firm proprietary, and firm broker/dealer account type. For example, trades entered by retail investors or institutional investors, such as hedge funds, are categorized as customer orders. Professional customer are highly active customers, whereas firms are member of the Options Clearing Corporation (OCC) like Morgan Stanley or Goldman Sachs that trade on behalf of their own accounts or for another broker/dealer who is not a member of the exchange. Market maker trades are reported indirectly under the market-clearing condition, as they usually take the opposite side of customer and firm trades.[5] We remove any duplicate entries and aggregate buy and sell volumes for each account type over opening and closing transactions. We end up with eight categories covering buy and sell volumes for each of the four trader types by option series and trading day.

2.2 Methodology to Infer the True Trade Side

Because evaluating the performance of trade classification algorithms requires information on the true side of the trade, we combine information from intraday transaction data and daily Open/Close data to arrive at such a benchmark. Our three Open/Close datasets are available on a daily level and cover trading volume at the ISE, the CBOE, and the GEMX, respectively. For that reason, we select all transactions in LiveVol that were executed at one of the three exchanges on days for which the total trading volume for an option series at the exchange equals either total customer sell volume or total customer buy volume in the Open/Close data. We take advantage of the fact that if there were only customer buy (sell) orders on a specific day for a given option series at one particular exchange, Open/Close data allows to classify all transactions in the LiveVol dataset on that day at the respective exchange as buy (sell) orders. We focus our analysis on trades between customers and market makers and assume that the customer is the party with a demand for options.[6] Therefore, we use the customer buy/sell indicator obtained from Open/Close data as the benchmark to empirically validate the accuracy of trade classification methods applied to intraday option transactions from LiveVol. We use the unique key specified by trade date, expiration date, strike price, option type, and root symbol of the underlying to match the samples. As a result,

[5] A detailed description of the ISE and GEMX Open/Close data is available at https://www.nasdaq.com/solutions/nasdaq-open-close-trade-profiles%3A-ise-and-gemx.

[6] Trades between customers and market makers are the most common trade constellation. They account for about 66% of trades at the ISE, 79% at the CBOE, and 81% at the GEMX. A more detailed description of trader type constellations is provided in Appendix A.2.

we obtain three matched samples, one for each option exchange, which contain price and quote data, along with a dummy variable indicating whether the trade was a customer buy or a sell. To analyze how classification accuracy depends on option characteristics and underlying information, we merge the data with OptionMetrics.

As discussed in the introduction, we follow a two-step research plan and use separate samples to, first, compare existing methods and develop our two new classification rules "in-sample", and, second, test our algorithms "out-of-sample". Our in-sample analyses are based on the matched ISE sample. The observation period for this dataset covers twelve years from May 2, 2005 to May 31, 2017. The starting date corresponds to the coverage of ISE Open/Close data going back to May 2005 and the end date is governed by the availability of the dataset to us when we performed our in-sample analysis. The matched ISE sample contains 49,203,747 option trades. In the second step, we test our newly developed classification algorithms out-of-sample on the CBOE and GEMX datasets after registering the results from the first step with the OSF (see footnote 3). The CBOE dataset begins on January 1, 2011 as there was a structural change in how CBOE Open/Close data is constructed at the beginning of 2011, and ends on October 31, 2017. The GEMX dataset covers the period from August 5, 2013 to June 30, 2021, with the starting date corresponding to the availability of GEMX Open/Close data.[7] The matched CBOE and GEMX samples contains 37,155,412 and 40,332,234 option trades, respectively

Table 1 reports summary statistics on our matched samples in Panels A.1 to C.1 and compares them with the full samples of all option trades on these three exchanges in Panels A.2 to C.2. On average, the matched samples cover between 10,000 and 14,000 option series per day that are written on 1,100 to 1,400 underlyings per day. The about 1.5-1.7 trades per option day result in a total of more than 125 million observations over the entire sample period. Comparing the number of unique options and underlyings per day shows that our matched samples represent on average between 36% and 60% of the option series and between 56% and 93% of the underlyings trading on the three exchanges. Overall, our matched samples cover around 15% of all ISE trades,

[7] We also obtained data for the ISE from June 1, 2017 until June 30, 2021 after uploading the results from the in-sample analysis to OSF. The results are qualitatively similar to our other out-of-sample tests. We therefore omit detailed results from the paper, but show time series of success rates in our analysis of changes over time (see Figure 4).

12% of CBOE trades, and 23% of trades at the GEMX. Our trade-side indicator based on the Open/Close data shows that buyer-initiated trades account for 47.5% of trades in the matched ISE sample, 45.0% in the CBOE sample, and 47.0% in the GEMX sample. As the probability of observing only buy trades or only sell trades decreases with an increasing number of trades, the number of trades per option day is lower and the time between two trades is higher in our matched samples compared to their full sample equivalents. Because tick tests depend on the information from preceding or succeeding trades as a precise signal for the fair option price, our results might therefore underestimate their performance. Looking at the average trade size and time to maturity shows that the trades in our matched sample are on average smaller and time to maturity is longer compared to the full samples, indicating an overall lower liquidity.

To address the point that liquidity might affect the performance of trade classification rules, we also test the performance of our trade size rule on a more liquid sample using an analogous GEMX Open/Close intraday dataset that is available on a 10-minute frequency. The observation period from August 5, 2013 to June 30, 2021 is identical to the GEMX end-of-day data. We match this data with option transactions from LiveVol in a similar way as with the end-of-day Open/Close data in our main analyses. That means we select all transactions in LiveVol that were executed at the GEMX within a 10-minute interval for which the total trading volume of that option series equals either the total customer sell volume or the total customer buy volume in the GEMX Open/Close intraday data. Panel C.3 of Table 1 reports summary statistics on our matched sample based on the intraday Open/Close data. Comparing summary statistics with our matched sample using end-of-day data (see Panel C.1 in Table 1) shows that the matched intraday sample covers a broader cross-section of option trades with a higher average trading frequency, suggesting higher liquidity in the sample. Overall, this sample covers around 66% of all GEMX trades.

3 Comparison of Trade Classification Algorithms

3.1 Applying Stock Market Classification Rules to Options

We compare the ability of common stock trade classification rules to correctly infer option trade direction. Table 2 presents an overview of the most common algorithms and their description. From a methodological perspective, they can be divided into three groups. First, quote rules compare trade prices to bid and ask quotes at the time of a trade. If the trade occurs above the midpoint of the bid-ask spread, it is classified as buyer-initiated. Conversely, if the trade price is below the midspread, the trade is classified as seller-initiated. Trades that occur exactly at the midpoint cannot be classified. The quote rule can be applied both to the NBBO or the bid and ask prices quoted at the trading venue at the time of the trade. To reduce the number of unclassifiable trades, Muravyev and Ni (2020) suggest to first apply the quote rule on the NBBO and then on the quotes at the executing exchange to classify trades at the NBBO midpoint. If both midpoints coincide, midspread trades still cannot be classified.

Second, tick tests use changes in trade prices and look at previous trade prices to infer trade direction. If the trade occurs at a higher price than the previous one, it is classified as buyer-initiated. Conversely, if the trade price is below the previous one, it is classified as seller-initiated. If there is no price change between successive trades, the trade direction is inferred using the last price that differs from the current price. This corresponds to assigning the same trade direction as for the previous trade. An alternative classification approach to tick tests, which look back at previous trades, are reverse tick tests that use the next trade price to classify the current trade. If the next trade price that is different from the price of the trade being classified is below the current price, the trade is classified as buyer-initiated. Conversely, if the next distinguishable price is above the current price, the current trade is classified as seller-initiated. The tick test and reverse tick test can be applied using trade prices on all option exchanges or one specific exchange only. While trade price information collected across all exchanges are updated more frequently and provide the most recent trade price, the price changes on the exchange on which the trade is executed might be more relevant to assess trades on this specific exchange.

Third, hybrid methods combine quote and tick rules to various degrees. The most common methods are the algorithms by Lee and Ready (1991) and Ellis, Michaely, and O'Hara (2000). The LR approach uses the quote rule to classify all trades that do not occur at the midpoint and the tick test to classify midspread trades. The EMO method uses the quote rule to classify only at-quote trades, for which the trade price matches the bid or ask price, and the tick test for all others. Besides these two, other combinations of quote and tick rules are also possible.

We now evaluate the success rates of the existing trade classification algorithms by comparing their classification to the true direction of the trade inferred from Open/Close data. To make the performance of algorithms that are unable to completely classify all trades comparable, we assume unclassified trades to be correctly classified with a random probability of 50%. This affects quote rules only, as they are unable to classify midspread trades.

Table 3 presents the success rates of common trade classification rules and their variations when applied to option trades at the ISE. The performance of quote rules is higher when applied to NBBO quotes compared to ISE quotes. Around 8.6% of the trades occur at the midspread and consequently cannot be classified by the quote rule. Using ISE quotes to classify trades occurring at the NBBO midspread reduces the number of unclassifiable trades that must be randomly assigned to about 6%. The successive use of NBBO and ISE quotes improves the performance of the quote rule to 63.85%. In general, quote rules are clearly superior to tick tests in classifying option trades. Tick tests perform best when using trade price information across all exchanges compared to just from the ISE. The success rate of the tick test using ISE trades only is even lower than 50%, which makes it worse than a random assignment of buys and sells. Interestingly, reverse tick tests that use subsequent trade prices have a higher success rate than their counterparts that use preceding trade prices to infer trade direction. However, the success rate of 55.71% using information from all exchanges is lower than for quote rules.

The results on the relative performance of different quote and tick rule specifications also carry over to the hybrid methods. Consequently, applying LR and EMO algorithms to NBBO quotes and price information across all exchanges as well as using subsequent trade prices to infer trade direction yields higher success rates. Moreover, we find that the LR algorithm outperforms the

EMO rule as, in addition to midspread trades, the latter uses the tick test to a greater extent. However, the commonly used LR rule using the tick test to classify midspread trades is only able to classify 63.53% of trades correctly, which is worse than using the quote rule alone. This result is due to the poor performance of the tick test, which correctly classifies midspread trades with a probability of less than 50%. The highest performance of 63.92% is achieved by the combination of the quote rule successively applied to NBBO and ISE quotes and the reverse tick test to classify all remaining trades that occur at the midpoint of both NBBO and ISE. The combination of quote and reverse tick rules, which we will refer to as "reverse LR" hereafter, has not yet been considered in the literature. Overall, the accuracy of common stock trade classification algorithms is significantly lower in our option dataset compared to stock trades (see, e.g., Lee and Ready (1991) or Ellis, Michaely, and O'Hara (2000)).

The last two columns of Table 3 show that the weak performance is mainly driven by trades with trade sizes equal to either the bid quote size or the ask quote size at the ISE at the time of the trade. Panel A.1 in Table 1 shows that this is the case for a relatively large fraction of 22.3% of all trades. For them, average success rates of quote rules are only about 30%.

3.2 New Classification Rules for Options

In this section, we develop two new rules to sign option trades. As noted in the introduction and in Section 2.2, we develop the new rules using ISE trades only. In Section 4, we then test the new rules out-of-sample on the CBOE and GEMX datasets.

Trade size rule

We start with the hypothesis that the weak performance of existing trade classification methods for trades with a trade size equal to either the size of the ask or the bid quote is due to limit orders placed by sophisticated customers. Market makers that completely fill such limit orders trigger the trade execution with a trade size equal to the quote size. In such a situation, the customer buys at the prevailing bid and sells at the prevailing ask, leading to misclassifications by the original quote rule.

Based on this idea, we propose to classify trades for which the trade size is equal to the quoted bid size as customer buys and those with a trade size equal to the ask size as customer sells. This rule alone can classify 22.3% of all trades at the ISE, of which 79.92% are correctly classified (see Table 4). After applying this "trade size rule", the existing trade classification algorithms are applied to all other trades for which the trade size is not equal to one of the quote sizes (or for which it is equal to both the bid and the ask size). Panel A of Table 5 shows that this modification leads to a substantial improvement between 10.7% and 11.3% in the performance of the quote rule and combined methods and an improvement of 5.6% to 7.3% for the tick tests. The highest success rates can be achieved by our new trade size rule together with the quote rule first applied to NBBO and then to ISE quotes. It correctly classifies 74.79% when applied solely or in combination with the reverse tick rule (reverse LR) and 74.64% in combination with the regular tick rule (LR). The result, that neither the tick test nor the reverse tick test is able to improve the success rates of the quote rule, for which trades at the midspread are randomly assigned as buys and sells, leads us to conclude that classifications based on preceding and succeeding trade prices are less informative in option markets than equity markets. This can be explained by the fact that tick rules perform better in highly liquid markets with more recent trade prices that contain more up-to-date information.

Depth rule

Our second new rule makes use of this finding and we propose an alternative approach to classify midspread trades. We hypothesize that a larger bid or ask quoted size, i.e., a higher depth at the best bid or ask, indicates a higher liquidity similar to a tighter bid or ask quote. As a consequence, we classify midspread trades as buyer-initiated, if the ask size exceeds the bid size, and as seller-initiated, if the bid size is higher than the ask size. If the ask size matches the bid size, midspread trades still cannot be classified by this approach, and we use the reverse tick test to classify such trades. Applying our proposed "depth rule" after using the trade size rule and quote rules leads to a success rate for midspread trades of 63.87% at the ISE (see Table 4), leading to an overall improved performance by around 0.8%. Only a negligible portion of this improvement is attributed to the application of the reverse tick test. We coin this combination "depth rule + reverse LR" and summarize the results in the last four rows in Panel A of Table 5. The success rate of the quote rule applied to NBBO first and then to ISE quotes increases to 75.51%, followed by the NBBO quote rule with 75.36%. Because the performance exceeds that of LR algorithms, we conclude that our depth rule outperforms the tick test and the reverse tick test in classifying midspread trades.

Based on our findings so far, we recommend that researchers use our new trade size rule together with quote rules successively applied to NBBO and quotes on the trading venue. Quotes at the midpoint on both the NBBO and the exchange should be classified first with the depth rule and any remaining trades with the reverse tick test. Most importantly, the LR algorithm alone, which is heavily used in the literature (see, e.g., Pan and Poteshman (2006); Hu (2014); Easley, O'Hara, and Srinivas (1998)), does a poor job to identify buy and sell orders in option trade data.[8] Overall, the accuracy of all common classification algorithms to infer option trade direction can be significantly improved by our two new rules.

Subsample analyses

To further challenge our hypothesis and to validate the improvement by our new trade size rule, we conduct two subsample analyses. First, we evaluate the performance of our trade size rule separately for different locations of trade prices relative to the bid and ask quotes at the ISE. The results in Panel B of Table 5 show that the new rule works best for trades occurring at the ask or bid quote and improves the success to classify at-quote trades by up to 21%. Contrary, the trade size rule even deteriorates the performance to correctly classify outside-quote trades compared to the traditional trade classification approaches by up to 4.5%. These results are also in line with our hypothesis that market makers fill limit orders from customers at the limit price, set by the customer. Contrary, if the trade price is outside the bid-ask spread, this is an indication that a customer wanted to trade against a standing limit order of a market maker, but the size of the market maker's quote was not sufficient, leading to a further price deterioration.

As a second subsample analysis, we evaluate the performance separately for various trade size categories. Figure 1 shows average success rates for the different specifications of the quote, tick, LR, reverse LR, EMO, and depth rules after the trade size rule has been applied for different trade

[8] OptionMetrics recently started offering a product ("IvyDB Signed Volume") that provides buying and selling volume information. Their classification is also based on the LR algorithm (see OptionMetrics (2020)).

size bins.[9] The cutoffs for the bins are calculated as quintiles and are measured in number of contracts. We show the overall success rates of the classification algorithms using our trade size rule and also calculate the change in the success rates compared to the same algorithms not using the trade size rule. The results show that our new rule works best for small to medium-sized trades and even leads to a slight deterioration of the performance for the largest trade sizes. This finding is in line with the hypothesis that limit orders placed by customers are more likely to be smaller trades. In contrast, large trades for which the trade size is equal to the quote size are more likely to be market orders in which customers want to trade the full depth of the market maker's bid or ask quote.

Given the results of the two subsample analyses, it would be possible to further improve the methodology of applying the trade size rule. In additional results, which are not tabulated to conserve space, we find that not applying the new rule for very large trades and for trades outside the bid-ask spread leads to small additional improvements of up to 0.4%.[10] Due to the additional complexity, which is prone to a potential over-fitting regarding the cutoff between small and large trades, and also due to the results from the out-of-sample tests that show mixed results for these additional refinements, we recommend to apply our two new rules to all trades.

4 Out-of-Sample Tests

To ensure that our new rules are not developed and tested on the same dataset, we conduct out-of-sample tests and repeat our main analyses using two similarly matched samples of option trades at the CBOE and the GEMX.

[9] We employ all the rules from Table 5, assign them into five groups, and report average results for the rules within the same group (four quote rules, two tick tests, four specifications of the LR algorithm, four specifications of the reverse LR algorithm, and four specifications of our depth rule used in combination with the reverse LR algorithm). In addition, we consider four variations of the original EMO rule (see Table 3).

[10] These results are available in the document uploaded to OSF (see footnote 3).

4.1 Testing Existing Trade Classification Rules

Putting the out-of-sample results for the new rules into perspective, we first evaluate the success of the existing trade classification algorithms for these samples and compare their performances with that for the ISE sample. Tables 6 and 7 present the results for the matched CBOE and GEMX samples, respectively. Overall, the success rate of existing trade classification algorithms for trades at the CBOE is comparable in magnitude to the ISE sample, although it is slightly lower for all methods by on average 2.8%. Moreover, the relative performance of the trade classification rules is qualitatively similar to that observed for the ISE. Namely, tick tests perform best when using most current price information across all exchanges and reverse tick tests based on subsequent prices dominate their counterparts based on preceding ones. Again, tick tests perform significantly worse than quote rules and are only able to correctly classify slightly more than 50% of option trades, which is not much better than a random allocation of buys and sells. For this reason, the LR algorithm outperforms the EMO rule as the former uses tick tests to a smaller extent.

In striking contrast to the results so far, quote rules and the combined methods that rely on them do not work at all for trades at the GEMX. Their success rates are all well below 50%, making a purely random assignment superior to applying these classification rules. An explanation for their poor performance is that for more than 50% of the trades at the GEMX, the trade size is equal to the quote size (see Panels C.1 and C.2 of Table 1), which indicates that quote sizes are small and potentially posted by customers and not by market makers. As discussed in the introduction, such a trading behavior is in line with the maker-taker pricing model of the GEMX that uses fee-based incentives to promote limit orders by customers.[11]

The performance of tick tests at the GEMX is qualitatively similar to that observed for the ISE and CBOE, classifying slightly more than 50% of option trades correctly. Consistent with the ISE and CBOE samples, tick tests perform best when using most current price information across all exchanges and reverse tick tests based on subsequent prices dominate their counterparts based on preceding ones. Notwithstanding the underperformance of quote rules, the LR algorithm performs

[11]In contrast to GEMX, CBOE uses the pay-for-order-flow model, that incentivies brokers to route their market orders to the exchange, and the ISE employs a mixed pricing model featuring combinations of the two incentive structures. For a discussion of the different pricing models, see, e.g., Battalio, Shkilko, and Van Ness (2016).

better than the EMO rule, just as in the ISE and CBOE samples. This is because tick tests actually only perform better than quote rules for at-quote trades. Both the LR algorithm and the EMO rule classify at-quote trades using the quote rule, but differ in the classification of trades occurring inside or outside the spread, for which quote rules perform still better than tick tests.

In contrast to the ISE sample, the quote rule performs better when applied to quotes from the trading venue, CBOE or GEMX, compared to the NBBO. Likewise, the successive use of trading venue and NBBO quotes yields slightly higher success rates when applied to quotes from the CBOE or GEMX first. A more detailed analysis shows that consistently across all three exchanges, for trades inside the NBBO or exchange bid-ask spread, quotes from the trading venue are better suited to classify trades than the NBBO. For trades at or outside the bid-ask spread, there is no clear pattern. We conjecture that the reason for the superior performance of using quotes from the trading venue for inside-spread trades are hidden orders and/or price improvements, for which information from the exchange's quotes might be more relevant. This finding could be used to further refine the application of quote rules, but we refrain from that as, first, potential improvements are very small, and second, we cannot test these refinements out-of-sample.

The last two columns of Tables 6 and 7 reveal that the overall weak performance of existing trade classification rules is again mainly driven by trades with trade sizes equal to either the bid quote size or the ask quote size. For them, average success rates of quote rules are only about 19% at the CBOE, and only about 9% at the GEMX. They account for 56% of the matched GEMX sample, as compared to 22% in the ISE sample and 14% in the CBOE sample, resulting in the extremely low performance at GEMX. Taking this into account, the success rate for all other trades at the GEMX is around 60%, which is above the 50% threshold and more comparable to the CBOE and ISE samples.

4.2 Testing New Classification Rules

Trade size rule

Because our trade size rule addresses exactly the trades with trade sizes equal to either the ask or the bid quote size, the poor performance of existing methods for these trades combined with the high prevalence of such trades on the GEMX exchange is a promising indication that our new rule also works in out-of-sample tests. We find that applying our trade size rule is successful for 85.47% of the trades on which it can be applied at the CBOE and in 92.75% of these cases at the GEMX, which is even better than for the ISE (see Table 4). While the proportion of trades that can be classified by our trade size rule is smaller in the CBOE sample, it is much higher in the GEMX sample. We therefore expect the overall improvement due to our trade size rule to be greater at the GEMX.

We now reevaluate the performance of the competing methods after our new trade size rule has been applied. That is, we use the existing trade classification rules to sign all other trades, for which the trade size is not equal to the quote size or for which the bid and ask quote sizes coincide. Panel A of Tables 8 and 9 present the results for the matched CBOE and GEMX samples, respectively. They reveal that our newly proposed trade size rule leads to a substantial improvement for both samples. For option trades at the CBOE, the classification success of the quote rules and combined methods is 9% higher on average compared to using the existing classification rules alone, and between 4.4% and 6% higher for the tick tests. The highest success rate of 72.40% is achieved by the trade size rule in combination with the quote rule first applied to CBOE quotes and then to the NBBO. In combination with the tick test or reverse tick test for the LR algorithm, we are able to correctly classify 72.12% and 72.39% of the option trades, respectively.

As expected, the improvement is even more pronounced in the GEMX sample, with a performance increase of 46.4% on average for the quote rules and combined methods, as well as 22.7% to 27.8% for the tick tests. The highest success rate of 80.48% is achieved by the reverse LR algorithm based on GEMX quotes after the trade size rule has been applied, closely followed by the reverse LR algorithm based on GEMX and NBBO quotes with 80.46% and the quote rule using GEMX and NBBO quotes alone with 80.38%.

Applying our trade size rule to the matched intraday sample at the GEMX in Panel A of Table 10 shows that our trade size rule also works in a more liquid environment. As expected, the improvements due to our new rules are somewhat smaller on average. Nevertheless, the trade size rule yields a substantial improvement in the classification success by on average 38% across the different specifications.

Depth rule

The fact that neither the tick test nor the reverse tick test is able to improve the success rates of the quote rules for trades at the CBOE and the GEMX reveals the shortcoming of the existing methods to classify midspread trades. We therefore test the success of our second new rule, the depth rule, to overcome these difficulties by classifying midspread trades based on the relative comparison of ask and bid quote sizes at the trading venue at the time of the trade. The results are summarized in the last four rows in Panel A of Tables 8 and 9 for the CBOE and GEMX samples, respectively. Applying our depth rule after using the trade size rule and the quote rule and classifying the very small number of midspread trades that cannot be signed by our depth rule using the reverse tick test yields an additional improvement of 1.21% on average at the CBOE, and 0.56% at the GEMX, respectively. It raises the highest success rate for CBOE trades to 73.37% when applied in combination with the quote rule that first uses CBOE quotes and then the NBBO, as well as 80.86% for GEMX trades when applied in combination with the quote rule that solely uses GEMX quotes. Finally, the results from the intraday sample at the GEMX in the last four rows of Panel A of Table 10 show a similar improvement for the depth rule than in the daily sample, although the overall performance of the best combination of rules is with 65.2% somewhat lower than in the intraday sample. Overall, these results confirm that our depth rule outperforms the standard and reverse tick test in classifying midspread trades.

Subsample analyses

Finally, we repeat the two subsample analyses from Section 3.2 to further validate the improvements by our trade size rule.

In the first subsample analysis, we evaluate the performance of our trade size rule separately for different locations of the trade prices relative to the quotes on the exchange. The results in Panel B of Tables 8 to 10 show that the trade size rule works best for trades occurring at the ask or bid quote. Our new rule improves the performance in classifying at-quote trades by up to 32% at the

CBOE and 59% at the GEMX, whereas it diminishes the success to correctly classify outside-quote trades by up to 2.2% at the CBOE and 2.8% at the GEMX as compared to the traditional trade classification approaches.

In the second subsample analysis, we evaluate the performance separately for various trade size categories. The results in Figures 2 and 3 show again that the trade size rule works best for small to medium-sized trades. In addition, it is successful for trades of all sizes on the CBOE and the GEMX, although to a smaller extent for larger trades.

In summary, the out-of-sample tests support the idea that the weak performance of traditional classification methods, which were developed to sign stock trades, is caused by misclassified limit orders submitted by customers. Our results show that substantial improvements are possible when our new trade size and depth rules are applied. Overall, both rules applied in combination with the existing algorithms achieve a success rate of over 73% and 80% for trades in the daily samples from the CBOE and the GEMX, respectively, which is much closer to the classification precision for stock trades in the literature (see, e.g., Chakrabarty, Li, Nguyen, and Van Ness (2007)). The overall lower success rates in the intraday GEMX sample of about 65%, which are improved by on average 38% compared to not applying our new rules, show that trade classification is most difficult in a liquid market that is based on a maker-taker pricing model.

5 Classification Accuracy in the Cross-Section and Over Time

To verify the superior performance of our new rules when applied in different situations, we perform a battery of sample splits and analyze their performance over time. Following Savickas and Wilson (2003), we analyze sample splits based on option characteristics such as option and security type, time to maturity, and moneyness. Because the tick rule might be more problematic if there is a long time period between trades, we perform a sample split based on the time between trades. To conserve space, we compute average success rates for the different specifications of the quote, tick, LR, reverse LR, EMO, and depth rules (see footnote 9).

We present the results for the ISE, the CBOE, and the GEMX in Tables 11, 12, and 13, respectively. The tables show that classification accuracies are between 1% and 2% higher for calls compared to puts, which is partly driven by a higher improvement due to our trade size rule. Comparing the classification precision of options written on common stocks, index options, and options written on other underlyings (mainly ETFs), we find lower success rates for index options, which is consistent with Savickas and Wilson (2003). Interestingly, the improvements due to our trade size rule are particularly high for index options at the CBOE.

Comparing options with different maturities and moneyness, we find that our trade size rule achieves the highest improvements when applied to options with long maturities and deep-out-of-the money options. When we look at the performance of the original algorithms, not applying our trade size rule, i.e., subtracting the improvement in parentheses from the rule's performance, we find that the original rules perform particularly poor for those options. Therefore, our new trade size rule resolves this weakness of the existing trade classification algorithms and leads to the highest improvement where it is most needed.

The results on the sample splits regarding the time between trades confirm that tick rules have problems classifying trades when there is a long time between two consecutive trades. For example, in the ISE and CBOE samples, the performance advantage of the quote rules over the tick rules is about 10% for the lowest quintile and increases to on average about 15% for the three quintiles with the longest time between trades. Interestingly, after a critical threshold of the time between trades is exceeded, the performance difference no longer increases.

To analyze the performance improvements over time, Panel A of Figure 4 shows for each of the exchanges the percentage of trades for which our trade size rule can be applied as the trade size equals the quote size. This figure confirms that the percentage of applicable trades strongly varies over time. We conjecture that the drivers of these variations are changes in the fee structure and the relative positioning of the exchange in the competitive trading landscape. As the fee structure evolves over time with multiple fee changes occurring each year, and as we do not have access to a comprehensive database of fee changes and their announcements, it is challenging to trace back the reasons for all the ups and downs of this time series.[12] We therefore exemplarily look at two large

[12]Moreover, as argued by Asçioglu, Holowczak, Louton, and Saraoglu (2017), some option trading firms might incorporate fee changes only with a delay, because their order routing decisions are based on an ex-post analysis of

changes in the percentage of applicable trades for which we found corresponding changes in the trading process. First, at the end of January 2015 at the CBOE, the percentage of trades for which the quote size equals the trade size goes down from about 15% to just about 3%. An explanation for this change might be a reduction of rebates and an increase in the volumes that are necessary to be entitled to get those rebates.[13] Similarly, the percentage of applicable trades drops at the beginning of April 2017 from about 80% to just about 20-30% for GEMX. This change corresponds to a migration to a new trading platform, which followed the acquisition of the exchange by Nasdaq and was accompanied by a number of changes in the fee structure and the trading process.[14]

Panel B of Figure 4 shows that changes in the number of applicable trades directly transfer to the improvements due to the trade size rule. Most importantly, the impact of these changes on the final success rates, which we present in Panel C of Figure 4, are relatively small. This result shows that our new rules not only improve the average accuracy of trade classification rules, but also reduce variations of the success rates over time and in the cross section.

Summarizing the results from the sample splits and the time series analysis, we find that in all subsamples and for all existing trade classification algorithms, improvements due to the application of our new trade size rule are positive. The relative performance ranking for the classification methods is similar across all subsamples, i.e., applying our depth rule after using the trade size rule and quote rule and classifying the remaining midspread trades using the reverse tick test yields the highest classification success followed by the reverse LR and quote rule. Standard LR algorithms, for which tick rules are based on the preceding trade, are always worse than their "reverse" counterparts for which tick rules are based on the succeeding trade. The EMO rule is inferior to the LR algorithm, as it uses the tick rule to a greater extent. Using tick rules alone is always the worst choice. Most importantly, in contrast to standard and reverse tick tests, our newly proposed depth rule leads to a significant improvement compared to using the quote rule alone,

actually charged fees.

[13] The corresponding fee change was filed on January 14, 2015 and published on January 26, 2015 on the SEC's website: `https://www.sec.gov/rules/sro/cboe/cboearchive/cboearchive2015.shtml`.

[14] See `http://www.nasdaqtrader.com/MicroNews.aspx?id=OTA2017-19` in connection with multiple announcements on the SEC's website in the first months of 2017: `https://www.sec.gov/rules/sro/isegemini.htm`. An example of a change that might influence the number of applicable trades is an improvement for Preferred Market Makers (PMMs) at the expense of other customers in the allocation of orders, see `https://www.sec.gov/rules/sro/isegemini/2017/34-80239.pdf`.

pointing to its superior performance to sign midspread trades that quote rules cannot classify. The time series analysis reinforces that the large differences in the performance of the existing stock classification rules over time and across different trading venues can be cured to a large extent by applying our newly developed classification rules.

6 Predicting stock returns

To put the relevance of correct trade classification in a broader context, we evaluate the predictability of future stock returns based on option order imbalances when applying different trade classification rules. We find that the higher accuracy of our new rules translates to higher excess returns of a long-short strategy, thus reinforcing the finding from Hu (2014) that option order flow contains valuable information about the underlying stocks.

The literature recognizes two channels through which option order flow can affect the underlying stock prices. First, option trading can contain information about the fundamental values of the underlying stocks when informed traders resort to option markets to exploit their private information due to the higher leverage, lower transaction costs, or fewer short selling restrictions of option markets (Easley, O'Hara, and Srinivas (1998)). Second, even without private information, market makers' delta hedging activities can cause price pressure on the underlying stocks. Market makers usually take the opposite side in option trades and accommodate customer order imbalances. Therefore, market makers' positions often deviate substantially from the desired level and they engage in delta hedging by trading the underlying stocks (Hu (2014)). Both of these channels lead to higher (lower) average stock returns on the next day when a stock's option order imbalance is positive (negative).

We follow Hu (2014) and calculate the option order imbalance (OOI) by aggregating the signed option trading volume for each stock i on day t, weighted by the delta exposure of each option contract and scaled by the number of shares outstanding:

$$OOI_{i,t} = \frac{\sum_{j=1}^{n_{i,t}} 100\mathrm{Dir}_{i,t,j} \cdot \mathrm{delta}_{i,t,j} \cdot \mathrm{size}_{i,t,j}}{\mathrm{Num\_shares\_outstanding}_{i,t}}, \tag{1}$$

where $n_{i,t}$ is the number of option trades in stock i on day t, $size_{i,t,j}$ denotes the trade size, and $Dir_{i,t,j}$ is a dummy variable equal to one (negative one) if option trade j is initiated by a buyer (seller) according to a trade signing algorithm. First, we apply a standard version of the LR algorithm that classifies trades according to the quote rule based on NBBO quotes and uses the tick test (across all exchanges) to classify all remaining midspread trades. Second, we implement our new classification rules and use the trade size rule together with the quote rules successively applied to the NBBO and the quotes on the trading venue. Quotes at the midpoint are classified first with the depth rule and any remaining trades with the reverse tick test. $delta_{i,t,j}$ is the option price's sensitivity to the underlying stock price and captures the hedge ratio. The numerator thus expresses the directional trading intention about the underlying stock. Following Hu (2014), we scale the measure by the number of common shares outstanding.

Our analysis is based on option transaction data from LiveVol from January 1, 2004 to June 30, 2021. Compared to our previous analysis, we now extend our sample to include trades on all U.S. options exchanges. In addition to the minimal filters described in Section 2.1, we follow Hu (2014) and focus on options on common stocks only, and exclude options that expire within ten calendar days, options with zero strike prices, and those at the market open (the first 15 minutes) or the market close (the last five minutes). We take the option delta and the number of common shares outstanding from OptionMetrics. Our sample for the option order imbalance construction covers almost 1.8 billion option trades. Finally, we obtain daily returns on the underlying stocks from the Center for Research in Security Prices (CRSP).

Panel A of Table 14 presents summary statistics of the data set. Our sample contains between 1,466 and 3,416 stocks per day, with an average of 2,209. The number of trades is on average 406,224 per day with a daily trading volume of 5.4 million contracts. Our sample thus contains a larger number of option trades and firms per day compared to the sample of Hu (2014), which is due to the extension of the sample period to more than 17 years and our minimal list of filters.[15] The last three rows of Panel A of Table 14 present the time series averages of cross-sectional statistics for the main variables of our portfolio analysis. On days when no option on the underlying stock

[15]The sample period of Hu (2014) goes from April 2008 to August 2010. It covers on average 1,670 stocks and 273,102 trades per day and the mean daily trade volume is 4.8 million contracts.

is traded, the order imbalance is set to zero. Overall, we find that the option market is relatively balanced with a mean imbalance of -0.04 bp and -0.03 bp, pointing to a slightly negative option net demand from non-market-makers. The standard deviation is smaller for the measure based on our new rules. We conjecture that the reason is their lower classification error.

Next, following Hu (2014), we perform portfolio sorts to assess the stock return predictability based on the option order imbalances comparing existing and new trade classification rules. For this purpose, we form daily quintile portfolios based on our two order imbalance measures. We then implement a long-short investment strategy that buys the stocks in the quintile with the most positive order imbalance and sells the stocks in the quintile with the most negative order imbalance. The portfolios are rebalanced every day at the market close based on the current day's option order imbalance.

Panel B of Table 14 shows the average returns from the order imbalance strategies. The returns generally increase across the quintile portfolios. An outlier is the comparatively high return of the middle portfolio 3 that typically includes the stocks for which the order imbalance is 0.[16] This pattern is similar to observations made in Hu (2014) and is likely due to the on average smaller size of the stocks in this portfolio. For the order imbalance variables based on the standard LR algorithm (our new classification rules), the sell portfolio has an average return of 3.06 bp (2.50 bp) per day, and the buy portfolio has an average return of 7.27 bp (7.82 bp). The daily excess return for the long-short portfolio is 4.21 bp for the existing LR algorithm and 5.33 bp for our new rules, which amounts to 10.6% and 13.4% annually. With t-statistics above 10, the excess returns are statistically significant at the 1% level before and after controlling for the Fama and French (1993) risk factors as well as the four-factor model of Carhart (1997). Importantly, the predictive power of the order imbalance strategy is higher when using our newly proposed trade size and depth rules. The difference of 1.11 bp per day is statistically significant at the 5% level and corresponds to a relative improvement of 26.6% compared to the excess return using the standard LR. Interestingly, the annual Sharpe ratio of the long-short portfolio increases even more strongly from 2.65 to 4.07 when using our new rules. Our results thus reinforce the finding from Hu (2014) that option order

[16] Note that due to the large number of stocks for which order imbalance is exactly 0, this portfolio contains on average about 42% of the stocks.

imbalance predicts future stock returns and show that the higher accuracy of our new rules leads to higher excess returns.

7 Conclusion

This study presents a comprehensive comparison of common trade classification algorithms including the quote rule, the tick test, the Lee and Ready algorithm, and Ellis, Michaely, and O'Hara method and evaluates their performance to infer options' trade direction. We use a matched sample of LiveVol intraday transactions and Open/Close data from the ISE, the CBOE, and the GEMX. Employing our novel matching methodology, we can observe the true trade directions for trades for which there were only customer buy or only customer sell trades on a specific day for a given option series at the specific exchange.

We find that the success of the common stock trade classification algorithms to correctly infer the direction of option trades is considerably lower than for stocks. Our results support the hypothesis that the poor performance of traditional classification approaches is mainly driven by sophisticated customers who implement their trading strategies via limit orders, particularly for small trades that are not outside of the bid-ask spread. Our new classification methodology corrects for these problems and strongly improves existing methods. As a second new rule, we propose a method to classify midspread trades based on the relative comparison of bid and ask quoted depth. Using our new methodology allows to correctly classify between 65% and 81% of option trades in our sample, which is between 10% to 47% higher compared to the rules that are currently used in the literature. Applying our new rules to construct a long-short trading strategy for stocks based on option order imbalance increases annual excess returns from 10.6% to 13.4% and Sharpe ratios from 2.65 to 4.07.

The importance of correct trade classification is highlighted in several papers that point to biases in microstructure research due to incorrectly signed trades. For example, Finucane (2000) find biases in signed volume and effective spreads for both the tick test and LR algorithm and Chakrabarty, Li, Nguyen, and Van Ness (2007) reveal that errors in trade side classification can result in substantially biased estimates of effective spreads and price impacts of trades for LR and EMO rules. Similarly, Savickas and Wilson (2003) show that all common methods perform poorly at estimating effective spreads for options. Our findings also have implications for studies that analyze the impact of market design on trading costs and market quality. To calculate measures of total trading costs, these studies typically add exchange-specific rebates and fees to the bid-ask spread under the assumption that customers submit market orders (see, e.g., Battalio, Shkilko, and Van Ness (2016); Anand, Hua, and McCormick (2016)). Our results indicate that this assumption often does not hold and, on top of that, its violation is strongly related to the market design.

Our results are important for the trade-off, which data source to use to quantify dealer positions and customer demand. Some studies use Open/Close data directly to avoid employing trade classification rules, for which the performance in option markets was not clear, yet (see, e.g., Gârleanu, Pedersen, and Poteshman (2009); Christoffersen, Goyenko, Jacobs, and Karoui (2018); Ni, Pearson, Poteshman, and White (2021)). The disadvantage of this approach is that data have to be purchased for each option exchange separately. Additionally, Open/Close data sets are not available for the majority of exchanges and typically are only sourced from the ISE and/or the CBOE. Whereas the market share of these two exchanges was relatively high in the first decade of the century, it gradually declined in parallel to the increase in the total number of option exchanges to 16. As a result, relative market shares are only about 6.6% for the ISE and 10.9% for the CBOE in 2021.[17] In contrast, trade data is available in a combined dataset for all exchanges from LiveVol. Our results suggest that trade classification algorithms using our new rules applied to all option trades, with their performance well above 70%, might lead to a more complete picture than just using the "true" trade direction for a small minority of the trading volume.

[17] See p. 54-56 of sifma's Market Structure Compendium for 2021, available at `https://www.sifma.org/wp-content/uploads/2022/03/SIFMA-Insights-Market-Structure-Compendium-March-2022.pdf`.

References

Aitken, Michael and Alex Frino, 1996, The accuracy of the tick test: Evidence from the Australian stock exchange, *Journal of Banking & Finance* 20, 1715–1729.

Anand, Amber, Jian Hua, and Tim McCormick, 2016, Make-take structure and market quality: Evidence from the US options markets, *Management Science* 62, 3271–3290.

Asçioglu, Asli, Richard Holowczak, David Louton, and Hakan Saraoglu, 2017, The evolution of market share among the US options market platforms, *The Quarterly Review of Economics and Finance* 64, 196–214.

Battalio, Robert, Andriy Shkilko, and Robert Van Ness, 2016, To pay or be paid? The impact of taker fees and order flow inducements on trading costs in US options markets, *Journal of Financial and Quantitative Analysis* 51, 1637–1662.

Cao, Charles, Zhiwu Chen, and John M Griffin, 2005, Informational content of option volume prior to takeovers, *The Journal of Business* 78, 1073–1109.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chakrabarty, Bidisha, Bingguang Li, Vanthuan Nguyen, and Robert A Van Ness, 2007, Trade classification algorithms for electronic communications network trades, *Journal of Banking & Finance* 31, 3806–3821.

Chan, Kalok, Y Peter Chung, and Wai-Ming Fong, 2002, The informational role of stock and option volume, *The Review of Financial Studies* 15, 1049–1075.

Christoffersen, Peter, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui, 2018, Illiquidity premia in the equity options market, *The Review of Financial Studies* 31, 811–851.

Easley, David, Marcos Lopez de Prado, and Maureen O'Hara, 2016, Discerning information from trade data, *Journal of Financial Economics* 120, 269–285.

Easley, David, Maureen O'Hara, and Pulle Subrahmanya Srinivas, 1998, Option volume and stock prices: Evidence on where informed traders trade, *The Journal of Finance* 53, 431–465.

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.

Fama, Eugene F and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of financial economics* 33, 3–56.

Finucane, Thomas J, 2000, A direct test of methods for inferring trade direction from intra-day data, *Journal of Financial and Quantitative Analysis* 35, 553–576.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2005, Limit order book as a market for liquidity, *The review of financial studies* 18, 1171–1217.

Gârleanu, Nicolae, Lasse Heje Pedersen, and Allen M Poteshman, 2009, Demand-based option pricing, *Review of Financial Studies* 22, 4259–4299.

Goettler, Ronald L, Christine A Parlour, and Uday Rajan, 2005, Equilibrium in a dynamic limit order market, *The Journal of Finance* 60, 2149–2192.

Goyenko, Ruslan, Chayawat Ornthanalai, and Shengzhe Tang, 2015, Options illiquidity: determinants and implications for stock returns, *Rotman School of Management Working Paper.*

Goyenko, R. and C. Zhang, 2021, Demand Pressures and Option Returns, Working Paper.

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

Lee, Charles MC and Balkrishna Radhakrishna, 2000, Inferring investor behavior: Evidence from TORQ data, *Journal of Financial Markets* 3, 83–111.

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

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

Muravyev, Dmitriy and Xuechuan Charles Ni, 2020, Why do option returns change sign from day to night?, *Journal of Financial Economics* 136, 219–238.

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

Newey, W and K West, 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–708.

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.

OptionMetrics, 2020, Assessing Option Demand from Signed Volume Order Flow, Research Note.

Pan, Jun and Allen M Poteshman, 2006, The information in option volume for future stock prices, *The Review of Financial Studies* 19, 871–908.

Pöppe, Thomas, Sebastian Moos, and Dirk Schiereck, 2016, The sensitivity of VPIN to the choice of trade classification algorithm, *Journal of Banking & Finance* 73, 165–181.

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

ISE

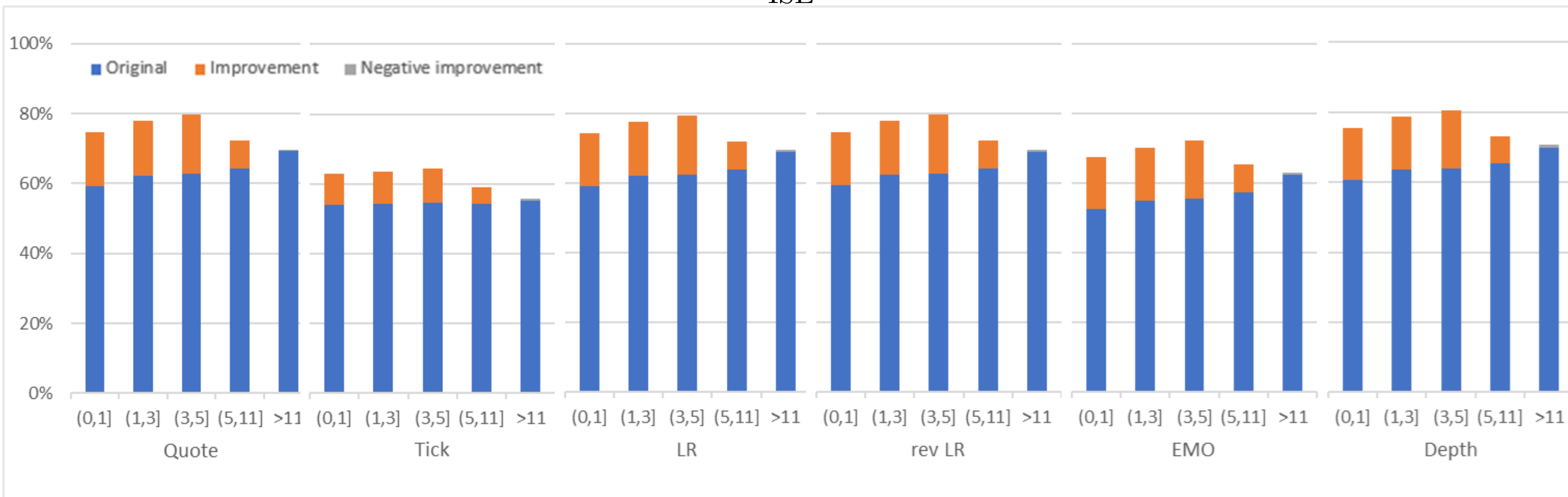


Figure 1: **Success rates of trade size rule for different trade size categories (ISE)**
This figure presents average success rates of competing trade classification methods after the trade size rule has been applied for different trade size bins. The trade size is measured in number of contracts and the cutoffs for the bins are calculated as quintiles. We take the average success of the rules across their variations, i.e., four quote rules, two tick tests, four specifications of the LR algorithm, the reverse LR algorithm, the EMO rule, as well as four specifications of our depth rule used in combination with the reverse LR algorithm (see footnote 9). The blue bar shows the success rate of the existing classification method, the orange bar represents the improvements due to the trade size rule, and the gray bar indicates a negative improvement.

CBOE

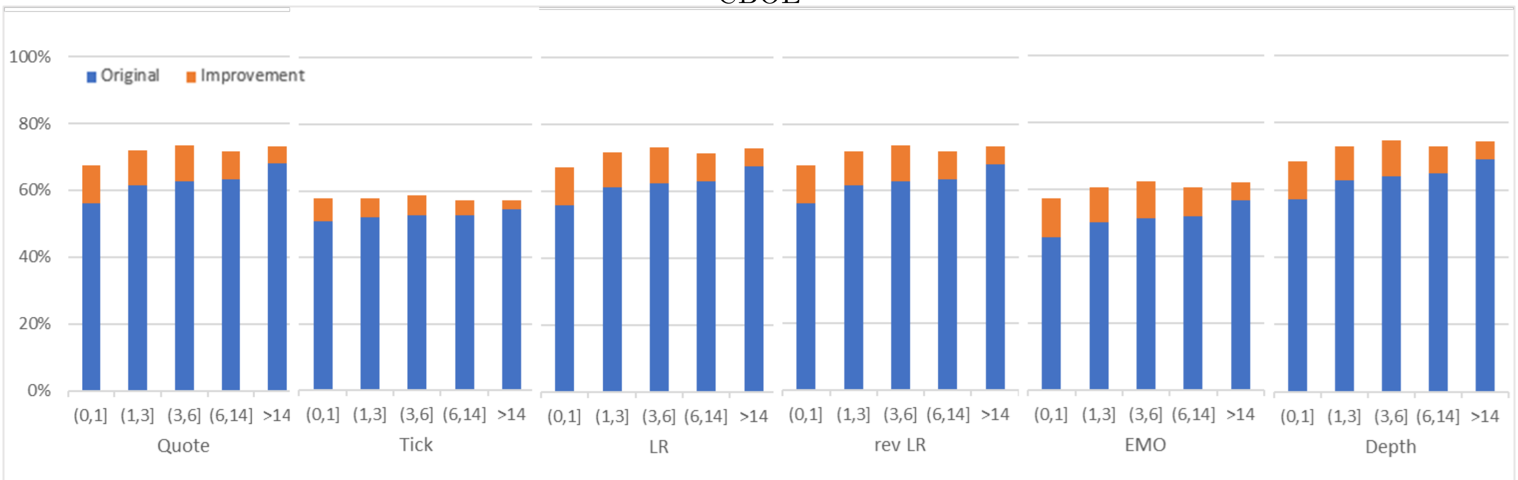
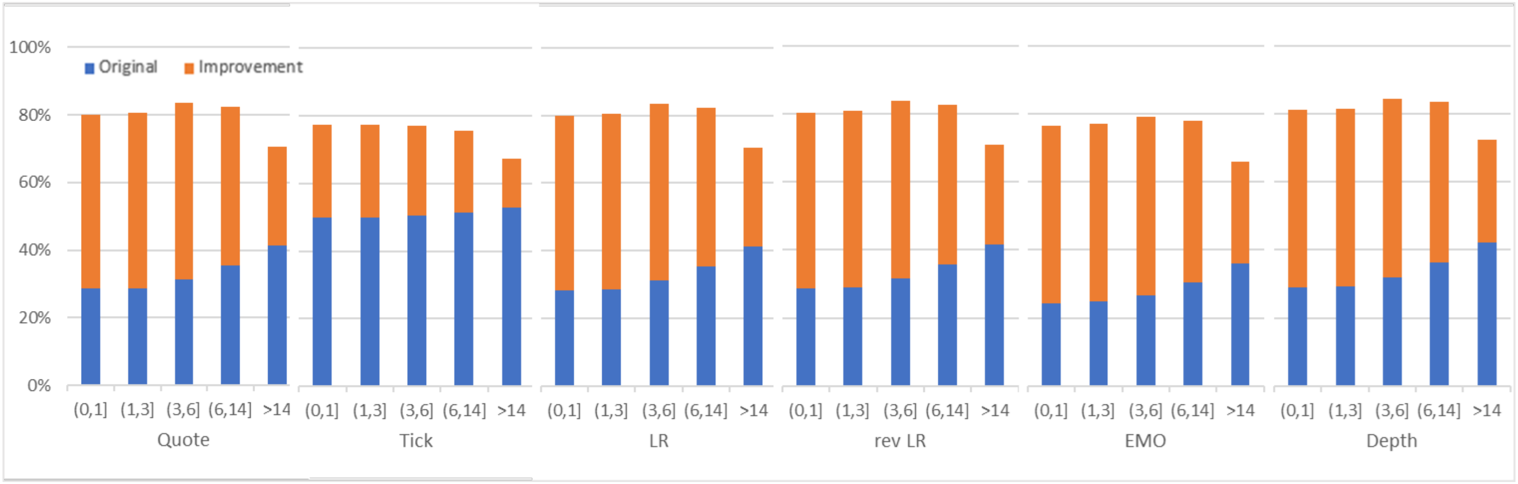


Figure 2: **Success rates of trade size rule for different trade size categories (CBOE)**
This figure presents average success rates of competing trade classification methods after the trade size rule has been applied for different trade size bins. The trade size is measured in number of contracts and the cutoffs for the bins are calculated as quintiles. We take the average success of the rules across their variations, i.e., four quote rules, two tick tests, four specifications of the LR algorithm, the reverse LR algorithm, the EMO rule, as well as four specifications of our depth rule used in combination with the reverse LR algorithm (see footnote 9). The blue bar shows the success rate of the existing classification method, while the orange bar indicates the improvements due to the trade size rule.

GEMX end-of-day



GEMX intraday

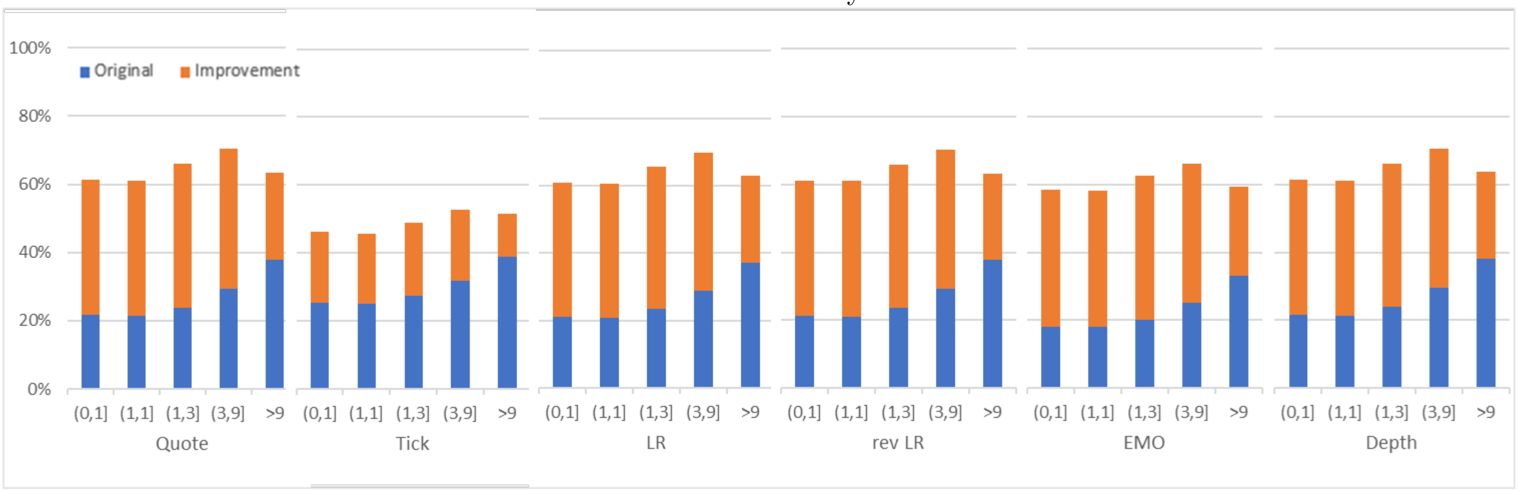


Figure 3: **Success rates of trade size rule for different trade size categories (GEMX)**
This figure presents average success rates of competing trade classification methods after the trade size rule has been applied for different trade size bins. The trade size is measured in number of contracts and the cutoffs for the bins are calculated as quintiles. We take the average success of the rules across their variations, i.e., four quote rules, two tick tests, four specifications of the LR algorithm, the reverse LR algorithm, the EMO rule, as well as four specifications of our depth rule used in combination with the reverse LR algorithm (see footnote 9). The blue bar shows the success rate of the existing classification method, while the orange bar indicates the improvements due to the trade size rule.

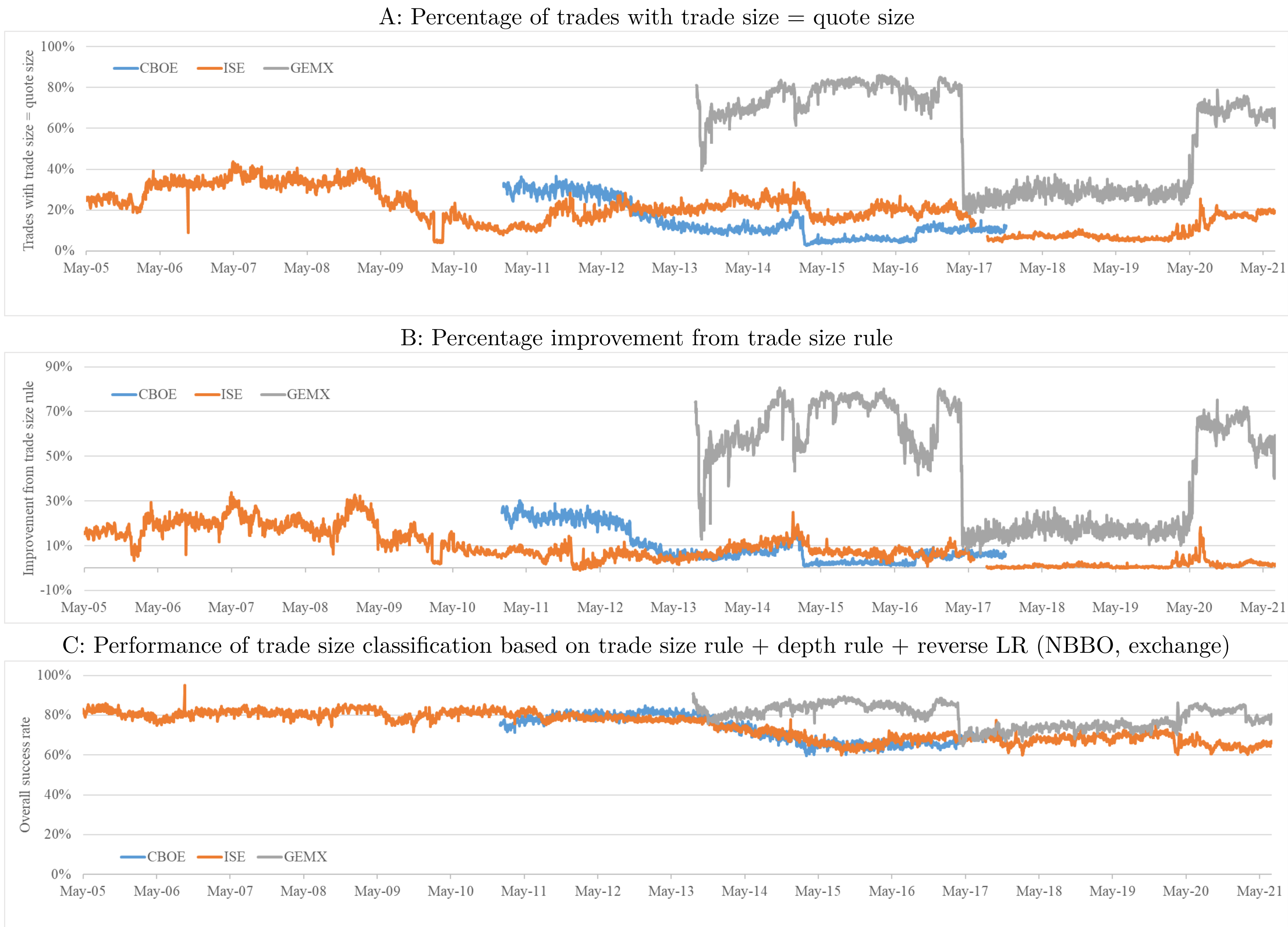


Figure 4: **Time variation of classification accuracy and improvements due to the trade size rule**
Panels A depicts the percentage of trades for which the trade size equals the quote size, and, therefore, the trade size rule can be applied. Panel B shows the percentage improvements due to our trade size rule for the best combination of rules from the in-sample analysis, i.e., applying the trade size rule, followed by the quote rule first on the NBBO and then on the quotes of the trading venue, followed by the depth rule and by the reverse tick test. Panel C presents the overall accuracy of this combination of rules. The blue line shows data for the CBOE, the orange line represents the ISE, and the gray line is for the GEMX.

Table 1: **Summary statistics**

This table shows descriptive statistics for our samples of LiveVol trade data matched with ISE Open/Close data (Panel A), CBOE Open/Close data (Panel B), and GEMX end of day as well as intraday Open/Close data (Panel C). The subpanels A.1, B.1, C.1 and C.3 provide statistics for our matched samples, whereas subpanels A.2, B.2 and C.2 compare them to all ISE, CBOE and GEMX trades from the full LiveVol dataset. The observation periods are May 2, 2005 to May 31, 2017 in Panel A, January 1, 2011 to October 31, 2017 in Panel B, and August 5, 2013 to June 30, 2021 in Panel C. We report the number of unique option series and underlyings per day, the number of trades per option-day as well as summary statistics on trade size measured in number of contracts, time between trades in hours, moneyness (i.e., underlying (strike) price relative to strike (underlying) price for call (put) options), and time to maturity in days. For each of these quantities, we report the mean, standard deviation, as well as the 5%, 50% (median), and 95% quantile. The total numbers of observations (N) are given in the top left corner of each panel. We also report the proportion of buy orders in the matched samples according to the Open/Close indicator, as well as the proportion of trades for which the trade size equals either the bid or ask quote size, as these trades are in the focus of one of our proposed rules.

Panel A: ISE

A.1: Matched ISE trades

N= 49,203,747	Mean	Std	5%	Median	95%
Option series per day	9,885.22	3,119.45	5,632	9,420	15,838
Underlyings per day	1,236.56	225.19	751	1,282	1,522
Trades per option day	1.64	2.49	1	1	4
Time between trades (in hours)	36.12	160.51	0.00	0.59	149.57
Trade size (# contracts)	13.62	77.75	1	4	50
Days to maturity	107.29	150.08	2	46	459
Moneyness	0.99	2.45	0.69	0.97	1.29
Trades with trade size = quote size	22.28%				
Buy trades	47.46%				

A.2: All trades recorded at the ISE

N= 337,234,107	Mean	Std	5%	Median	95%
Option series per day	22,388.14	4,773.00	14,192	22,579	29,792
Underlyings per day	1,468.24	293.97	778	1,535	1,795
Trades per option day	4.95	18.02	1	2	17
Time between trades (in hours)	9.70	85.65	0.00	0.01	24.80
Trade size (# contracts)	23.47	289.92	1	5	72
Days to maturity	62.21	106.58	1	29	236
Moneyness	0.98	5.92	0.76	0.98	1.14
Trades with trade size = quote size	25.58%				

Table 1 continued

Panel B: CBOE

B.1: Matched CBOE trades

N= 37,155,412	Mean	Std	5%	Median	95%
Option series per day	12,555.00	2,743.31	8,321	12,354	17,267
Underlyings per day	1,149.23	103.68	984	1,150	1,326
Trades per option day	1.72	3.07	1	1	4
Time between trades (in hours)	39.71	184.41	0.00	0.56	166.70
Trade size (# contracts)	18.14	223.24	1	5	50
Days to maturity	98.08	138.55	2	44	400
Moneyness	0.98	3.74	0.71	0.97	1.21
Trades with trade size = quote size	13.97%				
Buy trades	45.00%				

B.2: All trades recorded at the CBOE

N= 301,865,970	Mean	Std	5%	Median	95%
Option series per day	34,407.30	6,697.85	23,814.5	34,221	45,516
Underlyings per day	2,053.69	164.20	1,793	2,055	2,313
Trades per option day	5.10	20.59	1	1	17
Time between trades (in hours)	10.30	94.64	0.00	0.02	25.01
Trade size (# contracts)	24.07	258.02	1	5	84
Days to maturity	49.17	92.82	1	20	201
Moneyness	0.98	4.63	0.79	0.98	1.10
Trades with trade size = quote size	16.09%				

Table 1 continued

Panel C: GEMX

C.1: Matched GEMX trades (end-of-day Open/Close data)

N= 40,332,234	Mean	Std	5%	Median	95%
Option series per day	13,973.89	10,244.91	3,964	11,343	38,010
Underlyings per day	1,397.33	427.85	746	1,423	2,208
Trades per option day	1.47	1.94	1	1	3
Time between trades (in hours)	32.73	155.53	0.01	0.79	141.92
Trade size (# contracts)	8.82	34.22	1	2	30
Days to maturity	88.45	141.33	1	31	402
Moneyness	0.99	7.48	0.64	0.96	1.23
Trades with trade size = quote size	55.58%				
Buy trades	47.00%				

C.2: All trades recorded at the GEMX

N= 179,140,309	Mean	Std	5%	Median	95%
Option series per day	23,179.66	17,027.01	7,197	18,356	65,435
Underlying per day	1,509.20	411.58	898	1,539	2,247
Trades per option day	3.88	17.83	1	1	12
Time between trades (in hours)	10.16	88.50	0.00	0.04	26.47
Trade size (# contracts)	8.92	38.20	1	2	31
Days to maturity	42.64	95.98	0	10	200
Moneyness	0.97	5.77	0.72	0.98	1.10
Trades with trade size = quote size	51.68%				

C.3 Matched GEMX trades (intraday Open/Close data)

N= 118,837,177	Mean	Std	5%	Median	95%
Option series per day	19,596.84	15,907.49	5,045	15,051	58,312
Underlying per day	1,420.17	423.90	797	1,444	2,212
Trades per option day	3.05	7.56	1	1	10
Time between trades (in hours)	11.91	92.92	0.00	0.71	48.08
Trade size (# contracts)	8.03	33.76	1	2	29
Days to maturity	49.13	104.06	0	14	226
Moneyness	0.98	5.26	0.72	0.98	1.12
Trades with trade size = quote size	54.83%				
Buy trades	41.48%				

Table 2: **Description of common trade classification rules**

This table presents the commonly used trade classification algorithms in the microstructure literature. LR refers to the Lee and Ready (1991) algorithm and EMO refers to the algorithm introduced by Ellis, Michaely, and O'Hara (2000).

Rule	Description
Quote rule	Classifies a trade as a buy (sell) if its trade price is above (below) the midpoint of the bid and ask spread. Trades executed at the midspread are not classified.
Tick test	Classifies a trade as a buy (sell) if its trade price is above (below) the closest different price of a previous trade.
Reverse tick test	Classifies a trade as a buy (sell) if its trade price is above (below) the closest different price of a following trade.
LR	Classifies a trade as a buy (sell) if its price is above (below) the midpoint (quote rule), and uses the tick test to classify midspread trades.
EMO	Classifies a trade as a buy (sell) if the trade takes place at the ask (bid) quote, and uses the tick test to classify all other trades.

Table 3: **Success rates of common trade classification rules (ISE)**
This table presents the success rates of common trade classification rules and their variations when applied to option trades. Trades that cannot be classified by the respective rule are randomly assigned as buy or sell. The success rates of the combined methods refer to their application using trade prices across all exchanges. We report the percentage of unclassified trades, the overall success rate, and the success rate for trades for which the trade size equals the ask size or the bid size.

		% correctly classified		
	% not classified	all trades	ask size	bid size
Quote rule (NBBO)	8.58	63.69	31.83	30.90
Quote rule (ISE)	9.07	62.65	29.68	28.59
Quote rule (NBBO, ISE)	6.03	63.85	31.28	30.23
Quote rule (ISE, NBBO)	6.03	63.38	30.01	28.87
Tick test (ISE)	0.00	49.11	44.30	46.34
Tick test (all exchanges)	0.00	53.22	47.03	47.16
Reverse tick test (ISE)	0.00	52.31	47.21	56.81
Reverse tick test (all exchanges)	0.00	55.71	53.10	56.33
LR (NBBO)	0.00	63.53	32.00	30.96
LR (NBBO, ISE)	0.00	63.72	31.44	30.30
LR (ISE)	0.00	62.53	29.89	28.66
LR(ISE, NBBO)	0.00	63.25	30.17	28.94
Reverse LR (NBBO)	0.00	63.80	32.12	31.55
Reverse LR (NBBO, ISE)	0.00	63.92	31.49	30.72
Reverse LR (ISE)	0.00	62.83	29.96	29.18
Reverse LR (ISE, NBBO)	0.00	63.45	30.22	29.36
EMO (NBBO)	0.00	57.05	31.35	32.63
EMO (NBBO, ISE)	0.00	56.99	30.85	32.11
EMO (ISE)	0.00	55.36	30.22	31.47
EMO (ISE, NBBO)	0.00	56.99	30.85	32.11
Reverse EMO (NBBO)	0.00	57.74	31.93	34.39
Reverse EMO (NBBO, ISE)	0.00	57.65	31.39	33.77
Reverse EMO (ISE)	0.00	56.03	23.00	33.24
Reverse EMO (ISE, NBBO)	0.00	57.65	31.39	33.75

Table 4: **Performance of our new classification rules**
This table summarizes the percentage of classifiable trades and the success rate to correctly classify those trades using our new trades size rule and depth rule for our matched samples. The numbers are given as a percentage.

		ISE	CBOE	GEMX	Intraday GEMX
Trade size rule	% classifiable trades	22.28	13.97	55.58	54.83
	success (classifiable trades)	79.92	85.47	92.27	77.51
Depth rule	% classifiable trades	5.01	5.38	0.99	0.89
	success (classifiable trades)	63.87	67.97	83.33	73.25

Table 5: **Success rates of trade size rule for all trades and by relative location to quote (ISE)**

This table presents the success rates of competing classification rules after the trade size rule has been applied. Panel A shows the overall performance for the full sample, while Panel B breaks down the results for different subsamples in terms of locations of the trade price relative to the bid and ask quotes at the ISE. The success rates are given as a percentage. The improvements due to the trade size rule are in parentheses.

	Panel A	Panel B			
	all trades	at quote	at mid	inside	outside
Quote rule (NBBO)	74.50	80.99	60.39	70.58	77.90
	(10.81)	(21.00)	(4.45)	(2.18)	(-4.35)
Quote rule (ISE)	73.95	80.90	55.29	70.90	80.09
	(11.31)	(21.27)	(5.29)	(2.84)	(-4.48)
Quote rule (NBBO, ISE)	74.79	80.90	60.39	71.29	80.12
	(10.95)	(21.25)	(4.45)	(2.22)	(-4.48)
Quote rule (ISE, NBBO)	74.61	80.90	60.39	70.90	80.09
	(11.24)	(21.27)	(4.45)	(2.84)	(-4.48)
Tick test (all exchanges)	60.53	68.07	53.22	54.56	67.11
	(7.31)	(13.20)	(4.90)	(2.12)	(-3.48)
Reverse tick test (all exchanges)	61.35	68.96	55.04	54.97	66.38
	(5.64)	(10.03)	(4.15)	(1.71)	(-3.23)
LR (NBBO)	74.32	80.98	58.55	70.50	79.43
	(10.78)	(20.99)	(4.14)	(2.18)	(-4.48)
LR (NBBO, ISE)	74.64	80.90	58.55	71.29	80.12
	(10.92)	(21.25)	(4.14)	(2.22)	(-4.48)
LR (ISE)	73.80	80.90	53.22	70.90	80.09
	(11.27)	(21.27)	(4.90)	(2.84)	(-4.48)
LR (ISE/NBBO)	74.46	80.90	58.55	70.90	80.09
	(11.21)	(21.27)	(4.14)	(2.84)	(-4.48)
Reverse LR (NBBO)	74.51	81.01	60.03	70.58	78.60
	(10.71)	(20.95)	(3.54)	(2.16)	(-4.42)
Reverse LR (NBBO, ISE)	74.79	80.90	60.03	71.29	80.12
	(10.87)	(21.25)	(3.54)	(2.22)	(-4.48)
Reverse LR (ISE)	74.04	80.90	55.04	70.90	80.09
	(11.21)	(21.27)	(4.15)	(2.84)	(-4.48)
Reverse LR (ISE/NBBO)	74.61	80.90	60.03	70.90	80.09
	(11.16)	(21.27)	(3.54)	(2.84)	(-4.48)
Depth rule + reverse LR (NBBO)	75.36	81.05	68.78	70.83	77.51
	(10.32)	(20.73)	(0.45)	(2.08)	(-4.31)
Depth rule + reverse LR (NBBO, ISE)	75.51	80.90	68.78	71.29	80.12
	(10.61)	(21.25)	(0.45)	(2.22)	(-4.48)
Depth rule + reverse LR (ISE)	74.95	80.90	66.11	70.90	80.09
	(10.89)	(21.27)	(0.23)	(2.84)	(-4.48)
Depth rule + reverse LR (ISE, NBBO)	75.33	80.90	68.78	70.90	80.09
	(10.91)	(21.27)	(0.45)	(2.84)	(-4.48)

Table 6: **Success rates of common trade classification rules (CBOE)**
This table presents the success rates of common trade classification rules and their variations when applied to option trades. Trades that cannot be classified by the respective rule are randomly assigned as buy or sell. The success rates of the combined methods refer to their application using trade prices across all exchanges. We report the percentage of unclassified trades, the overall success rate, and the success rate for trades for which the trade size equals the ask size or the bid size.

		% correctly classified		
	% not classified	all trades	ask size	bid size
Quote rule (NBBO)	9.10	60.78	16.52	21.83
Quote rule (CBOE)	9.70	62.45	16.59	21.79
Quote rule (NBBO, CBOE)	5.70	61.57	16.27	21.66
Quote rule (CBOE, NBBO)	5.70	63.10	16.63	21.86
Tick test (CBOE)	0.00	47.82	42.30	42.20
Tick test (all exchanges)	0.00	50.80	42.34	42.51
Reverse tick test (CBOE)	0.00	51.44	46.60	58.26
Reverse tick test (all exchanges)	0.00	54.09	51.02	57.15
LR (NBBO)	0.00	60.36	16.48	21.78
LR (NBBO, CBOE)	0.00	61.28	16.26	21.60
LR (CBOE)	0.00	62.03	16.57	21.72
LR (CBOE, NBBO)	0.00	62.81	16.62	21.80
Reverse LR (NBBO)	0.00	60.80	16.59	22.00
Reverse LR (NBBO, CBOE)	0.00	61.56	16.31	21.75
Reverse LR (CBOE)	0.00	62.52	16.63	21.92
Reverse LR (CBOE, NBBO)	0.00	63.09	16.67	21.94
EMO (NBBO)	0.00	51.48	15.88	20.81
EMO (NBBO, CBOE)	0.00	51.47	15.50	20.49
EMO (CBOE)	0.00	48.95	15.23	20.08
EMO (CBOE, NBBO)	0.00	51.47	15.50	20.49
Reverse EMO (NBBO)	0.00	52.89	16.50	21.60
Reverse EMO (NBBO, CBOE)	0.00	52.87	16.11	21.25
Reverse EMO (CBOE)	0.00	50.17	15.85	20.80
Reverse EMO (CBOE, NBBO)	0.00	52.87	16.11	21.25

Table 7: **Success rates of common trade classification rules (GEMX end-of-day)**
This table presents the success rates of common trade classification rules and their variations when applied to option trades. Trades that cannot be classified by the respective rule are randomly assigned as buy or sell. The success rates of the combined methods refer to their application using trade prices across all exchanges. We report the percentage of unclassified trades, the overall success rate, and the success rate for trades for which the trade size equals the ask size or the bid size.

		% correctly classified		
	% not classified	all trades	ask size	bid size
Quote rule (NBBO)	4.34	31.74	9.22	9.22
Quote rule (GEMX)	2.57	34.08	9.54	9.60
Quote rule (NBBO, GEMX)	1.21	32.54	9.18	9.19
Quote rule (GEMX, NBBO)	1.21	34.14	9.53	9.58
Tick test (GEMX)	0.00	43.58	40.74	45.14
Tick test (all exchanges)	0.00	44.79	43.24	42.02
Reverse tick test (GEMX)	0.00	51.85	41.62	63.89
Reverse tick test (all exchanges)	0.00	57.05	53.93	59.60
LR (NBBO)	0.00	31.35	9.20	9.21
LR (NBBO, GEMX)	0.00	32.42	9.17	9.18
LR (GEMX)	0.00	33.82	9.51	9.57
LR (GEMX, NBBO)	0.00	34.03	9.52	9.57
Reverse LR (NBBO)	0.00	32.01	9.28	9.30
Reverse LR (NBBO, GEMX)	0.00	32.62	9.19	9.21
Reverse LR (GEMX)	0.00	34.26	9.56	9.63
Reverse LR (GEMX, NBBO)	0.00	34.23	9.54	9.60
EMO (NBBO)	0.00	28.69	8.75	8.85
EMO (NBBO, GEMXE)	0.00	28.58	8.51	8.64
EMO (GEMX)	0.00	28.07	8.41	8.58
EMO (GEMX, NBBO)	0.00	28.58	8.51	8.64
Reverse EMO (NBBO)	0.00	31.68	9.33	9.47
Reverse EMO (NBBO, GEMX)	0.00	31.58	9.10	9.28
Reverse EMO (GEMX)	0.00	31.15	9.01	9.22
Reverse EMO (GEMX, NBBO)	0.00	31.58	9.10	9.28

Table 8: **Success rates of trade size rule for all trades and by relative location to quote (CBOE)**

This table presents the success rates of competing classification rules after the trade size rule has been applied. Panel A shows the overall performance for the full sample, while Panel B breaks down the results for different subsamples in terms of locations of the trade price relative to the bid and ask quotes at the CBOE. The success rates are given as a percentage. The improvements due to the trade size rule are in parentheses.

	Panel A	Panel B			
	all trades	at quote	at mid	inside	outside
Quote rule (NBBO)	70.09	76.40	55.79	69.13	75.69
	(9.31)	(31.74)	(0.51)	(0.15)	(-2.13)
Quote rule (CBOE)	71.76	76.35	50.61	72.92	77.26
	(9.31)	(31.90)	(0.61)	(0.05)	(-2.19)
Quote rule (NBBO, CBOE)	70.91	76.35	55.79	70.49	77.29
	(9.34)	(31.90)	(0.51)	(0.12)	(-2.19)
Quote rule (CBOE, NBBO)	72.40	76.35	55.79	72.92	77.26
	(9.30)	(31.90)	(0.51)	(0.05)	(-2.19)
Tick test (all exchanges)	56.81	70.62	46.27	51.95	62.80
	(6.01)	(19.76)	(0.63)	(0.38)	(-1.40)
Reverse tick test (all exchanges)	58.53	72.67	50.48	53.01	63.80
	(4.45)	(14.65)	(0.53)	(0.26)	(-1.75)
LR (NBBO)	69.68	76.43	52.76	68.91	76.90
	(9.32)	(31.72)	(0.53)	(0.16)	(-2.15)
LR (NBBO, CBOE)	70.63	76.35	52.76	70.49	77.29
	(9.35)	(31.90)	(0.53)	(0.12)	(-2.19)
LR (CBOE)	71.35	76.35	46.27	72.92	77.26
	(9.32)	(31.90)	(0.63)	(0.05)	(-2.19)
LR (CBOE, NBBO)	72.12	76.35	52.76	72.92	77.26
	(9.31)	(31.90)	(0.53)	(0.05)	(-2.19)
Reverse LR (NBBO)	70.09	76.42	55.55	69.15	76.10
	(9.30)	(31.71)	(0.47)	(0.15)	(-2.16)
Reverse LR (NBBO, CBOE)	70.89	76.35	55.55	70.49	77.29
	(9.34)	(31.90)	(0.47)	(0.12)	(-2.19)
Reverse LR (CBOE)	71.82	76.35	50.48	72.92	77.26
	(9.30)	(31.90)	(0.53)	(0.05)	(-2.19)
Reverse LR (CBOE, NBBO)	72.39	76.35	55.55	72.92	77.26
	(9.30)	(31.90)	(0.47)	(0.05)	(-2.19)
Depth rule + reverse LR (NBBO)	71.49	76.44	66.26	69.82	75.32
	(9.22)	(31.58)	(0.15)	(0.12)	(-2.13)
Depth rule + reverse LR (NBBO, CBOE)	71.88	76.35	66.26	70.49	77.29
	(9.31)	(31.90)	(0.15)	(0.12)	(-2.19)
Depth rule + reverse LR (CBOE)	73.28	76.35	66.27	72.92	77.26
	(9.26)	(31.90)	(0.03)	(0.05)	(-2.19)
Depth rule + reverse LR (CBOE, NBBO)	73.37	76.35	66.26	72.92	77.26
	(9.27)	(31.90)	(0.15)	(0.05)	(-2.19)

Table 9: **Success rates of trade size rule for all trades and by relative location to quote (GEMX)**

This table presents the success rates of competing classification rules after the trade size rule has been applied. Panel A shows the overall performance for the full sample, while Panel B breaks down the results for different subsamples in terms of locations of the trade price relative to the bid and ask quotes at the GEMX. The success rates are given as a percentage. The improvements due to the trade size rule are in parentheses.

	Panel A	Panel B			
	all trades	at quote	at mid	inside	outside
Quote rule (NBBO)	78.16	83.17	52.49	60.89	84.97
	(46.42)	(58.90)	(2.70)	(1.82)	(-2.68)
Quote rule (GEMX)	80.31	83.19	52.45	72.35	85.54
	(46.23)	(59.03)	(2.45)	(0.34)	(-2.81)
Quote rule (NBBO, GEMX)	78.98	83.19	52.49	65.05	85.73
	(46.44)	(59.03)	(2.70)	(1.41)	(-2.81)
Quote rule (GEMX, NBBO)	80.38	83.19	52.49	72.35	85.54
	(46.24)	(59.03)	(2.70)	(0.34)	(-2.81)
Tick test (all exchanges)	72.62	80.27	44.10	45.64	56.76
	(27.83)	(34.71)	(3.19)	(3.37)	(-0.53)
Reverse tick test (all exchanges)	77.18	82.60	57.61	57.68	62.84
	(20.13)	(25.29)	(1.83)	(1.73)	(-2.58)
LR (NBBO)	77.78	83.19	47.73	59.42	85.43
	(46.43)	(58.87)	(3.04)	(1.97)	(-2.76)
LR (NBBO, GEMX)	78.87	83.19	47.73	65.05	85.73
	(46.45)	(59.03)	(3.04)	(1.41)	(-2.81)
LR (GEMX)	80.06	83.19	44.10	72.35	85.54
	(46.24)	(59.03)	(3.19)	(0.34)	(-2.81)
LR (GEMX, NBBO)	80.27	83.19	47.73	72.35	85.54
	(46.24)	(59.03)	(3.04)	(0.34)	(-2.81)
Reverse LR (NBBO)	78.39	83.17	55.95	61.69	85.22
	(46.38)	(58.89)	(2.29)	(1.73)	(-2.72)
Reverse LR (NBBO, GEMX)	79.05	83.19	55.95	65.05	85.73
	(46.43)	(59.03)	(2.29)	(1.41)	(-2.81)
Reverse LR (GEMX)	80.48	83.19	57.61	72.35	85.54
	(46.21)	(59.03)	(1.83)	(0.34)	(-2.81)
Reverse LR (GEMX, NBBO)	80.46	83.19	55.95	72.35	85.54
	(46.23)	(59.03)	(2.29)	(0.34)	(-2.81)
Depth rule + reverse LR (NBBO)	79.16	83.17	68.20	64.33	84.80
	(46.21)	(58.79)	(0.92)	(1.40)	(-2.69)
Depth rule + reverse LR (NBBO, GEMX)	79.31	83.19	68.20	65.05	85.73
	(46.40)	(59.03)	(0.92)	(1.41)	(-2.81)
Depth rule + reverse LR (GEMX)	80.86	83.19	75.66	72.35	85.54
	(46.17)	(59.03)	(-0.28)	(0.34)	(-2.81)
Depth rule + reverse LR (GEMX, NBBO)	80.72	83.19	68.20	72.35	85.54
	(46.20)	(59.03)	(0.92)	(0.34)	(-2.81)

Table 10: **Success rates of trade size rule by relative location to quote (GEMX Intraday)**
This table presents the success rates of competing classification rules after the trade size rule has been applied for different subsamples in terms of locations of the trade price relative to the bid and ask quotes at the Gemini. The success rates are given as a percentage. The improvements due to the trade size rule are in parentheses.

	Panel A	Panel B			
	all trades	at quote	at mid	inside	outside
Quote rule (NBBO)	63.84	66.09	54.55	54.79	72.54
	(37.70)	(46.93)	(1.21)	(0.99)	(4.44)
Quote rule (GMX)	65.15	66.09	62.09	61.47	72.98
	(37.60)	(47.03)	(12.09)	(0.23)	(4.35)
Quote rule (NBBO, GMX)	64.01	66.09	54.58	55.76	73.14
	(37.80)	(47.03)	(1.21)	(1.14)	(4.34)
Quote rule (GMX, NBBO)	65.01	66.09	54.58	61.47	72.98
	(37.64)	(47.03)	(1.21)	(0.23)	(4.35)
Tick test (all exchanges)	60.38	65.65	40.17	39.42	54.13
	(23.20)	(28.30)	(3.06)	(2.96)	(3.45)
Reverse tick test (all exchanges)	64.99	68.54	51.40	50.72	58.19
	(15.58)	(19.18)	(1.64)	(1.22)	(0.75)
LR (NBBO)	62.86	66.10	43.96	50.54	72.96
	(37.85)	(46.92)	(2.89)	(1.69)	(4.37)
LR (NBBO, GMX)	63.77	66.09	43.96	55.76	73.14
	(37.84)	(47.03)	(2.89)	(1.14)	(4.34)
LR (GMX)	64.63	66.09	40.17	61.47	72.98
	(37.68)	(47.03)	(3.06)	(0.23)	(4.35)
LR (GMX/NBBO)	64.78	66.09	43.96	61.47	72.98
	(37.68)	(47.03)	(2.89)	(0.23)	(4.35)
Reverse LR (NBBO)	63.46	66.09	50.54	53.12	72.76
	(37.79)	(46.93)	(2.12)	(1.39)	(4.42)
Reverse LR (NBBO, GMX)	63.92	66.09	50.54	55.76	73.14
	(37.82)	(47.03)	(2.12)	(1.14)	(4.34)
Reverse LR (GMX)	64.94	66.09	51.40	61.47	72.98
	(37.65)	(47.03)	(1.64)	(0.23)	(4.35)
Reverse LR (GMX/NBBO)	64.93	66.09	50.54	61.47	72.98
	(37.66)	(47.03)	(2.12)	(0.23)	(4.35)
Depth rule + reverse LR (NBBO)	63.99	66.09	59.20	55.11	72.24
	(37.66)	(46.88)	(0.85)	(1.10)	(4.50)
Depth rule + reverse LR (NBBO, GMX)	64.11	66.09	59.20	55.76	73.14
	(37.79)	(47.03)	(0.85)	(1.14)	(4.34)
Depth rule + reverse LR (GMX)	65.20	66.09	63.57	61.47	72.98
	(37.61)	(47.03)	(-0.27)	(0.23)	(4.35)
Depth rule + reverse LR (GMX, NBBO)	65.11	66.09	59.20	61.47	72.98
	(37.63)	(47.03)	(0.85)	(0.23)	(4.35)

Table 11: **Success rates of trade size rule for different subsamples (ISE)**
This table presents the average success rate of competing classification methods across their variations after the trade size rule has been applied for different subsamples based on option characteristics such as option and security type, time to maturity in days, moneyness (i.e., underlying (strike) price relative to strike (underlying) price for call (put) options) and time from previous trade in seconds. The security type category "Others" contains options on ETFs, Mutual Funds, and American Depositary Receipts. The cutoffs for time from previous trade bins are calculated as quintiles. The success rates are given as a percentage. The improvements due to the trade size rule are in parentheses.

	Quote	Tick	LR	rev LR	EMO	Depth
All trades	74.47	60.94	74.30	74.48	67.40	75.29
	(11.08)	(6.48)	(11.05)	(10.99)	(10.81)	(10.68)
Option type						
Call options	75.36	61.67	75.21	75.39	68.54	76.18
	(12.01)	(6.96)	(11.99)	(11.93)	(11.74)	(11.65)
Put options	73.28	59.98	73.11	73.30	65.91	74.12
	(9.84)	(5.83)	(9.81)	(9.75)	(9.58)	(9.41)
Security type						
Stock options	74.36	61.24	74.21	74.39	67.60	75.19
	(11.49)	(6.66)	(11.46)	(11.40)	(11.22)	(11.11)
Index options	68.17	55.69	68.12	68.16	56.80	68.51
	(7.15)	(4.06)	(7.14)	(7.12)	(7.16)	(6.95)
Others	75.15	60.30	74.95	75.14	67.38	75.99
	(10.04)	(6.05)	(10.01)	(9.93)	(9.74)	(9.58)
Time to maturity						
ttm <= 1 month	73.78	60.49	73.64	73.76	66.68	74.66
	(9.75)	(5.89)	(9.72)	(9.65)	(9.53)	(9.27)
ttm (1-2] month	75.59	60.56	75.38	75.57	68.15	76.45
	(10.46)	(6.25)	(10.45)	(10.39)	(10.34)	(10.11)
ttm (2-3] month	76.79	61.30	76.59	76.75	69.18	77.62
	(11.61)	(6.71)	(11.59)	(11.55)	(11.50)	(11.32)
ttm (3-6] month	73.57	60.54	73.40	73.64	66.91	74.43
	(10.37)	(6.08)	(10.34)	(10.27)	(9.99)	(9.96)
ttm (6-12] month	76.67	62.00	76.50	76.67	68.71	77.38
	(14.42)	(8.20)	(14.39)	(14.36)	(14.27)	(14.13)
ttm > 12 month	75.18	64.14	75.13	75.26	68.36	75.68
	(17.22)	(9.29)	(17.18)	(17.15)	(16.80)	(16.95)

Table 11 continued

	Quote	Tick	LR	rev LR	EMO	Depth
Moneyness						
mny <=0.7	72.02	61.07	71.87	72.27	66.30	72.77
	(10.94)	(5.87)	(10.92)	(10.90)	(10.84)	(10.77)
mny (0.7-0.9]	75.50	61.22	75.32	75.50	68.63	76.19
	(9.54)	(5.40)	(9.50)	(9.48)	(9.42)	(9.29)
mny (0.9-1.1]	74.35	60.82	74.20	74.34	67.30	75.23
	(10.80)	(6.52)	(10.77)	(10.69)	(10.51)	(10.28)
mny (1.1-1.3]	74.47	60.83	74.28	74.56	66.43	75.41
	(14.23)	(8.10)	(14.20)	(14.16)	(13.72)	(13.95)
mny > 1.3	73.88	61.02	73.71	73.96	66.20	74.55
	(15.18)	(8.38)	(15.16)	(15.13)	(14.73)	(14.99)
Time from previous trade						
<= 6 seconds	66.15	55.77	65.92	65.99	60.03	66.90
	(1.86)	(0.75)	(1.77)	(1.76)	(1.45)	(1.48)
(6-660] seconds	72.50	60.90	72.20	72.40	65.33	73.35
	(11.37)	(7.00)	(11.35)	(11.27)	(11.12)	(10.94)
(660-5,843] seconds	77.53	63.16	77.39	77.53	71.07	78.36
	(12.74)	(7.84)	(12.73)	(12.63)	(12.54)	(12.25)
(5,843-86,532] seconds	78.45	62.94	78.38	78.63	71.02	79.35
	(13.97)	(8.11)	(13.96)	(13.90)	(13.78)	(13.61)
> 86,532 seconds	77.85	62.29	77.78	78.03	70.07	78.64
	(15.34)	(8.55)	(15.33)	(15.28)	(15.01)	(15.04)

Table 12: **Success rates of trade size rule for different subsamples (CBOE)**
This table presents the average success rate of competing classification methods across their variations after the trade size rule has been applied for different subsamples based on option characteristics such as option and security type, time to maturity in days, moneyness (i.e., underlying (strike) price relative to strike (underlying) price for call (put) options) and time from previous trade in seconds. The security type category "Others" contains options on ETFs, Mutual Funds, and American Depositary Receipts. The cutoffs for time from previous trade bins are calculated as quintiles. The success rates are given as a percentage. The improvements due to the trade size rule are in parentheses.

	Quote	Tick	LR	rev LR	EMO	Depth
All trades	71.29	57.67	70.94	71.30	60.31	72.51
	(9.32)	(5.23)	(9.32)	(9.31)	(9.48)	(9.26)
Option type						
Call options	71.86	58.03	71.48	71.87	60.67	73.11
	(9.59)	(5.40)	(9.60)	(9.58)	(9.75)	(9.53)
Put options	70.58	57.23	70.26	70.57	59.86	71.74
	(8.97)	(5.01)	(8.98)	(8.96)	(9.13)	(8.92)
Security type						
Stock options	71.43	57.88	71.11	71.46	60.70	72.68
	(9.25)	(5.23)	(9.26)	(9.24)	(9.43)	(9.19)
Index options	65.48	57.52	65.15	65.50	54.60	66.59
	(13.44)	(7.13)	(13.44)	(13.42)	(13.45)	(13.40)
Others	72.40	57.19	71.99	72.35	60.77	73.56
	(8.45)	(4.74)	(8.46)	(8.44)	(8.59)	(8.40)
Time to maturity						
ttm <= 1 month	70.84	57.29	70.51	70.79	60.14	72.09
	(8.59)	(4.85)	(8.60)	(8.58)	(8.75)	(8.53)
ttm (1-2] month	72.18	57.56	71.84	72.12	61.64	73.35
	(9.43)	(5.37)	(9.44)	(9.42)	(9.59)	(9.38)
ttm (2-3] month	72.95	58.28	72.59	72.93	61.56	74.20
	(10.55)	(5.92)	(10.56)	(10.54)	(10.70)	(10.49)
ttm (3-6] month	70.87	57.52	70.46	70.95	59.94	72.18
	(8.79)	(4.94)	(8.80)	(8.78)	(8.96)	(8.74)
ttm (6-12] month	73.34	58.34	73.06	73.33	60.41	74.35
	(11.51)	(6.39)	(11.52)	(11.50)	(11.65)	(11.46)
ttm > 12 month	69.31	59.27	69.19	69.37	58.94	70.02
	(11.79)	(6.34)	(11.80)	(11.77)	(11.91)	(11.75)

Table 12 continued

	Quote	Tick	LR	rev LR	EMO	Depth
Moneyness						
mny <=0.7	68.34	58.35	67.96	68.44	61.07	69.64
	(10.41)	(5.54)	(10.41)	(10.39)	(10.43)	(10.36)
mny (0.7-0.9]	72.07	58.60	71.65	72.07	61.97	73.22
	(9.44)	(5.20)	(9.45)	(9.43)	(9.56)	(9.39)
mny (0.9-1.1]	71.96	57.35	71.62	71.95	60.23	73.19
	(8.78)	(4.99)	(8.79)	(8.77)	(8.97)	(8.72)
mny (1.1-1.3]	67.51	57.45	67.26	67.60	57.48	68.69
	(11.52)	(6.44)	(11.53)	(11.51)	(11.60)	(11.47)
mny > 1.3	66.34	57.51	66.05	66.41	56.61	67.53
	(12.46)	(6.82)	(12.47)	(12.45)	(12.52)	(12.41)
Time from previous trade						
<= 3 seconds	64.32	55.45	64.04	64.15	54.99	65.22
	(4.51)	(2.13)	(4.51)	(4.51)	(4.69)	(4.47)
(3-592] seconds	71.41	57.20	70.85	71.21	60.08	72.60
	(9.88)	(5.77)	(9.89)	(9.87)	(10.05)	(9.82)
(592-5,670] seconds	74.80	58.57	74.42	74.81	63.38	76.15
	(9.24)	(5.48)	(9.25)	(9.23)	(9.40)	(9.18)
(5,670-87,261] seconds	73.91	58.77	73.61	74.07	62.60	75.29
	(10.63)	(6.05)	(10.64)	(10.61)	(10.79)	(10.57)
> 87,261 seconds	72.29	58.93	72.04	72.50	61.41	73.55
	(12.25)	(6.60)	(12.26)	(12.24)	(12.34)	(12.19)

Table 13: **Success rates of trade size rule for different subsamples (GEMX)**
This table presents the average success rate of competing classification methods across their variations after the trade size rule has been applied for different subsamples based on option characteristics such as option and security type, time to maturity in days, moneyness (i.e., underlying (strike) price relative to strike (underlying) price for call (put) options) and time from previous trade in seconds. The security type category "Others" contains options on ETFs, Mutual Funds, and American Depositary Receipts. The cutoffs for time from previous trade bins are calculated as quintiles. The success rates are given as a percentage. The improvements due to the trade size rule are in parentheses.

	Quote	Tick	LR	rev LR	EMO	Depth
All trades	79.15	74.90	79.25	79.59	75.24	80.01
	(46.36)	(23.98)	(46.34)	(46.31)	(46.76)	(46.25)
Option type						
Call options	79.79	75.55	79.58	79.94	75.61	80.36
	(47.57)	(50.97)	(47.58)	(47.55)	(48.00)	(47.49)
Put options	78.93	73.86	78.71	79.05	74.66	79.47
	(44.37)	(50.85)	(44.38)	(44.36)	(44.80)	(44.29)
Security type						
Stock options	79.47	75.55	79.27	79.61	75.33	80.02
	(47.91)	(50.61)	(47.92)	(47.89)	(48.38)	(47.82)
Index options	62.91	73.04	62.90	62.93	60.54	62.98
	(45.83)	(51.34)	(45.83)	(45.82)	(46.19)	(45.81)
Others	79.66	72.70	79.43	79.82	75.16	80.26
	(40.90)	(51.97)	(40.90)	(40.88)	(41.18)	(40.82)
Time to maturity						
ttm <= 1 month	80.74	75.36	80.54	80.88	76.48	81.33
	(46.88)	(24.29)	(46.89)	(46.86)	(47.31)	(46.79)
ttm (1-2] month	80.42	74.72	80.20	80.57	75.85	81.06
	(45.72)	(23.84)	(45.73)	(45.70)	(46.18)	(45.62)
ttm (2-3] month	79.67	74.65	79.46	79.80	74.96	80.26
	(46.03)	(23.98)	(46.04)	(46.02v	(46.50)	(45.94)
ttm (3-6] month	78.32	74.40	78.09	78.48	74.27	78.89
	(45.36)	(23.40)	(45.37)	(45.35)	(45.78)	(45.28)
ttm (6-12] month	79.03	74.99	78.87	79.13	74.33	79.48
	(47.45)	(24.76)	(47.46)	(47.43)	(47.96)	(47.37)
ttm > 12 month	79.16	76.47	79.05	79.21	75.78	79.40
	(50.41)	(25.73)	(50.42)	(50.40)	(50.78)	(50.37)

Table 13 continued

	Quote	Tick	LR	rev LR	EMO	Depth
Moneyness						
mny <=0.7	77.69	74.69	77.51	78.05	74.11	78.15
	(44.38)	(21.34)	(44.39)	(44.39)	(44.75)	(44.32)
mny (0.7-0.9]	78.50	74.36	78.31	79.16	74.31	79.02
	(43.97)	(21.96)	(43.98)	(43.91)	(44.41)	(43.89)
mny (0.9-1.1]	80.09	74.78	79.87	81.27	75.59	80.68
	(46.27)	(24.31)	(46.28)	(46.13)	(46.72)	(46.17)
mny (1.1-1.3]	79.08	76.55	78.84	80.40	75.72	79.64
	(52.41)	(28.10)	(52.42)	(52.26)	(52.72)	(52.33)
mny > 1.3	78.92	77.47	78.74	79.87	76.38	79.35
	(54.05)	(28.17)	(54.06)	(53.95)	(54.31)	(53.99)
Time from previous trade						
<= 304 seconds	76.86	73.01	76.66	76.95	72.42	77.28
	(40.75)	(21.68)	(40.75)	(40.73)	(41.21)	(40.66)
(304-1,450] seconds	79.01	75.06	78.82	79.15	74.75	79.57
	(46.17)	(24.09)	(46.18)	(46.15)	(46.63)	(46.08)
(1,450-5,890] seconds	79.74	75.02	79.56	79.88	75.39	80.36
	(46.50)	(24.11)	(46.52)	(46.48)	(46.98)	(46.41)
(5,890-81,588] seconds	80.37	75.52	80.12	80.55	76.08	80.99
	(48.43)	(24.84)	(48.44)	(48.41)	(48.85)	(48.35)
> 81,588 seconds	80.96	75.88	80.72	81.13	77.23	81.56
	(49.41)	(24.84)	(49.42)	(49.39)	(49.74)	(49.33)

Table 14: **Portfolio sorts based on option order imbalance**

This table provides the results from our investment analysis using information from option order imbalance. Panel A describes the summary statistics of the full options sample. We provide time series averages of cross-sectional statistics for the stock return, as well as the option order imbalance (OOI), applying trade signing algorithms based on the standard Lee and Ready (1991) rules (standard LR) and our new trade size and depth rules (new rules). Returns and order imbalances are reported in basis points. Panel B presents the average daily returns on the quintile portfolios as well as the excess returns from the long-short portfolios based on option order imbalance (OOI) calculated with the standard LR algorithm and our new trade signing algorithms. The excess returns are reported as raw return, Fama and French (1993) three-factor alpha (FF3), and Carhart (1997) four-factor alpha (FF4). All returns are reported in basis points. Sharpe is the annualized Sharpe ratio. The Newey and West (1987) *t*-statistics are reported in parentheses. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% level.

Panel A: Summary statistics of full option sample

Variable	Mean	Std.	Median	Min	Max
Number of firms per day	2,209	325	2,169	1,466	3,416
Daily number of trades	406,224				
Daily trade volume	5,396,297				
OOI (standard LR)	-0.04	5.64	0.00	-212.03	134.28
OOI (new rules)	-0.03	5.14	0.00	-186.99	128.39
Stock return	6.25	329.35	-1.18	-3495.57	6400.76

Panel B: Portfolio returns

Quintile	OOI (existing rules)	OOI (new rules)	Difference
Low - 1	3.06	2.50	
2	4.00	4.23	
3	7.68	7.75	
4	3.98	5.26	
High - 5	7.27	7.82	
5-1	4.21***	5.33***	1.11**
	(10.88)	(16.16)	(2.28)
FF3 alpha	4.25***	5.25***	0.99**
	(10.97)	(15.99)	(2.05)
FF4 alpha	4.25***	5.25***	0.99**
	(11.00)	(16.05)	(2.05)
Sharpe	2.65	4.07	

A Appendix

A.1 Transaction costs of limit orders

Our paper builds on the literature of optimal order submission that analyzes the trade-offs between submission of market orders and limit orders (see, e.g., Foucault, Kadan, and Kandel (2005); Goettler, Parlour, and Rajan (2005)). For example, limit orders allow customers to trade at more favorable prices, but come at the cost of delayed or no trade execution. Moreover, an exchange's pricing model, e.g., a maker-taker fee structure that offers rebates to those who provide liquidity, can incentivize customers to post limit orders. Therefore, one can differentiate between two different motivations for customers to submit limit orders. First, customers may have a demand for options and try to meet their demand at more favorable prices than the prevailing quotes by implementing their trading strategies through limit orders. Market makers that decide to fill such limit orders trigger the trade execution and provide liquidity to the customer. Second, customers may act as liquidity providers to other market participants and post limit orders to profit from the bid-ask spread.

To better understand the customers' incentives for using limit orders, we analyse the return of a customer trading strategy that buys or sells an option at the given trade price during the day and closes the position at the end-of-day price. Panel A in Table A2 shows for the matched ISE sample that over all customer trades, the average return of buys and sells is negative with an average of -4.0% and -3.7%. This implies that customer buy prices are higher and customer sell prices are lower than end-of-day prices. Therefore, customers, to meet their option demand, on average lose money to market makers who provide liquidity.

We now compare the return of buys and sells for all customer trades to trades with trade sizes equal to the quote size that we identify as customer limit orders. Panel B in Table A2 shows that for such limit orders, the average return of buys and sells is still negative at -0.6% and -2.3%, respectively. This suggests that sophisticated customers submitting limit orders trade on average at more favorable prices, i.e., lower buying prices and higher selling prices, as compared to the sample of all trades including both market orders and limit orders. However, customers do not profit from providing liquidity through limit orders. Instead, our results indicate that customers are on average the party with a demand for options and market makers provide liquidity to them. This holds independently of whether customers use market orders or limit orders to implement their trading strategy.

A.2 Extension to different trader types

We focus our analysis on trades between customers and market makers as, in line with previous literature, we assume that in such constellations the customer is the party with a demand for options and market makers take the opposite position in the option trade. This allows us to clearly identify the trade direction. In contrast, trades between two customers are more problematic when determining the trade direction through classification algorithms. We therefore evaluate the frequency of such trade constellations. For this analysis, we use a matched sample of LiveVol and Open/Close data on days when an option series is traded only once at the respective exchange, as on such days a clear identification of the involved parties is possible through the Open/Close trader type categorization, i.e., customer, professional customer, firm proprietary, and firm broker/dealer. Market maker trades are reported only indirectly under the market-clearing condition, as they usually take the opposite side of customer and firm trades.[18]

Table A3 shows that trades between customers and market makers are the most common trade constellation. They account for 66.3% of trades at the ISE, 78.6% at the CBOE, and 81.4% at the GEMX. For these trades, as well as for trades between customers and professional customers or firms (included in other), it can be assumed that customers have a demand for options and play the more active role in a trade. In the same spirit, market makers usually take the opposite side of customer and firm trades. Trades in which no customer or market maker is involved play only a minor role, accounting for less than 0.32%. In general, our newly proposed trade size rule is consistent with option trades involving a single customer, a single market maker, or both.

For trades between two participants of the same trader type category, the true incentives to

[18] After a structural change to the CBOE Open/Close data due to the migration to Bats technology in 2011, the CBOE database offers more granularity in terms of trade origin, explicitly including market maker accounts. A detailed description of the CBOE Open/Close data is available at https://datashop.cboe.com/cboe-options-open-close-volume-summary.

trade are not clear, as both parties might have a demand for options, but classification algorithms always assign a trade direction. Trades between two customers are less frequent and only account for 2.6% of trades at the ISE, 0.6% at the CBOE, and 1.9% at the GEMX. Trades between two market makers account for 18.2% of trades at the ISE, 11.4% at the CBOE, and 8.25% at the GEMX. Interestingly, the proportion of customer-customer trades at the GEMX is in the same range as at the other two exchanges, even though GEMX uses fee-based incentives to promote limit orders by customers. However, instead of other customers filling these customers' limit orders, market makers take the opposite side in the trade more often. Our trade size rule identifies such trades through the simple heuristic that the market maker fills the complete size of the limit order and for such trades assigns option demand to the customer who initially placed the limit order as the trade initiator.

Table A1: **Literature on trade classification algorithms**
This table provides an overview of the success rates of different classification algorithms from existing studies for different asset classes and time periods. We report success rates (in percent) from the quote rule, the tick test, the Lee and Ready algorithm (LR, 1991), and the Ellis, Michaely, and O'Hara rule (EMO, 2000).

Study	Data	Sample period	Quote	Tick	LR	EMO
U.S. Stock market						
Lee and Radhakrishna (2000)	NYSE	11/1990 - 01/1991			93.3	
Finucane (2000)	NYSE	11/1990 - 01/1991		83.0	84.4	
Ellis, Michaely, and O'Hara (2000)	NASDAQ	09/1996 - 09/1997	76.4	77.7	81.1	81.9
Chakrabarty, Li, Nguyen, and Van Ness (2007)	NASDAQ	04/2005 - 06/2005		75.4	74.4	75.8
International stock market						
Aitken and Frino (1996)	Australian Stock Exchange	06/1992 - 06/1994		74.4		
Pöppe, Moos, and Schiereck (2016)	Deutsche Boerse	01/2012 - 11/2012		82.0	89.6	90.4
Other asset classes						
Easley, de Prado, and O'Hara (2016)	CME e-mini S&P 500 futures	11/2010 - 11/2011		86.4		
	COMEX gold futures	11/2010 - 11/2011		79.0		
	NYMEX oil futures	11/2010 - 12/2011		67.2		
Savickas and Wilson (2003)	CBOE options	07/1995 - 12/1995	82.8	59.4	80.1	76.5

Table A2: **Return of customer transactions at the day of initiation**

This table provides summary statistics on the return of a customer trading strategy that buys or sells an option at the given trade price during the day and closes the position at the end-of-day mid price. The return of buys (sells) is calculated as the positive (negative) difference between the end-of-day price and the buy (sell) price, divided by the buy (sell) price. We use the Open/Close identifier to classify customer buy and sell trades. Panel A provides statistics for all trades in the matched ISE sample, while Panel B considers only trades with a trade size equal to the quote size. The return is given as a percentage.

	N	Mean	Std	5%	Median	95%
Panel A: All trades in matched ISE sample						
Buys	23,334,617	-4.01	105.55	-50.00	-1.89	22.68
Sells	25,836,449	-3.67	281.84	-30.83	-0.89	21.74
Panel B: Trades with trade size equal to quote size						
Buys	5,123,277	-0.57	165.33	-28.57	-0.64	25.00
Sells	5,825,560	-2.36	585.37	-27.05	0.00	25.00

Table A3: **Trader type distribution**
This table describes the frequency of different constellations of trader types involved in a trade in the Open/Close Data. The samples consists of days with single option trades at the respective exchange, the ISE, the CBOE or the GEMX. The number of observations in each sample is provided at the bottom. The category "other" includes professional customer, firm proprietary, and broker/dealer. The frequency is given as a percentage.

Trader type	ISE	CBOE	GEMX
Customer - Market maker	66.28	78.55	81.39
Customer - Other	3.19	2.52	2.22
Customer - Customer	2.58	0.63	1.90
Market maker - Other	7.30	4.71	4.66
Market maker - Market maker	18.19	11.40	8.25
Other (same type)	0.18	0.07	0.17
Other (different type)	0.14	0.06	0.07
Unclassified	2.13	2.07	1.34
#Observations	33,308,074	19,110,220	26,403,650