The Rise of Off-Exchange Trading and the Role of Big Data AI in Price Discovery and Volatility

Abstract

We examine intraday and trade-by-trade factors associated with the big rise in off-exchange trading volume, information share, and price discovery from 2008 to 2022. Off-exchange trading has nearly doubled both in terms of volume and information share in the last two decades to 41.62% and 18.67%, respectively. In trade-by-trade analysis after informed options trades, we observe a dramatic jump in the probability from about 40% to about 60% that the next stock trade occurs off-exchange. We theoretically model and empirically identify informed traders' cost-based choice of off-exchange stealth trading venues, where they can avoid high predicted quoted stock spreads (forecasted by options activity). Our results sharply contrast the conventional wisdom about off-exchange volume being largely uninformed. Informed traders use low-cost off-exchange venues around earnings announcements even more intensely, particularly in light of wider spreads in those periods. This relation persists despite headwinds from option expiration settlement rules that drive up on-exchange volume. Our findings hold across traditional and machine learning methods for identifying informed trader intensity and surge around earnings announcements. Off-exchange information has important regulatory implications as it predicts higher subsequent stock price volatility. Our results are strongest in the most recent sample from 2020-22. This pattern implies that the immediacy of off-exchange has improved sufficiently in recent years to levels acceptable to informed traders who want to strategically exploit the low cost stealth trading feature of off-exchange dark pools.

Keywords: Market Microstructure, Off-exchange regulation, Quoted Spreads, Implied Volatility, Machine Learning, Informed Trading Intensity, Liquidity Suppliers

1 Introduction

Off-exchange trading has nearly doubled both in terms of volume and information share in the last two decades to 41.62% and 18.67%, respectively, as shown in Figure 1.[1] In trade by trade analysis after informed options trades, we observe a dramatic jump in the probability from about 40% to about 60% that the next stock trade occurs off-exchange.[2] Regulators and several recent academic publications (Brogaard and Pan (2022); Buti, Rindi, and Werner (2017, 2022); Comerton-Forde and Putniņš (2015)) raise important off-exchange issues of transparency, fairness, investor protection, and illiquidity risk premium. Despite their assessment of the effects of dark trading on market quality, we have limited understanding of the informational and cross-market determinants of the day-to-day variation in off-exchange dark trading and their impacts on the price discovery and volatility. Because off-exchange trades are derivatively priced based on national best bid and offer (NBBO) from exchanges which are important contributors to price discovery and market quality (Brogaard, Ringgenberg, and Roesch (2023)), the information share of off-exchange trades and their increasing role in price discovery is an important economic question and a mystery that our paper sets out to resolve. We provide novel evidence regarding off-exchange price discovery which we posit as an innovative form of stealth trading (Chakravarty (2001)) in the less transparent off-exchange trading environment (Topbas and Ye (2024)).[3] We are the first to ask whether the rise of off-exchange trading and its day-to-day variation in the pecking order of trading venue choices can be attributed to innovative low-cost order execution mechanisms and cross-market interaction between information and liquidity.

Figure 2 provides a summary of our stylized model and serves as a roadmap for our hypotheses and analysis. Our model extends Budish, Cramton, and Shim (2015) to include an additional

[1] We include all the trades marked as "D" from the DTAQ dataset at intra-day frequency, which contains dark pool trades and internalized orders. For sub-periods used by Topbas and Ye (2024) and Ernst et al. (2021), they report off-exchange market share similar to ours. Some papers use FINRA data at weekly frequency such as Brogaard and Pan (2022) and they document a growth from 7.5% in 2008 to 12.6% in 2015 for the dark pool subset.

[2] See Figure 4 for details.

[3] The importance of stealth trading is also highlighted by stock exchange's own follow-on attempts to launch their dark pools (Brugler and Comerton-Forde (2022)) and hidden orders (Bartlett and O'Hara (2024)).

options trading venue. Informed traders choose to trade options and stocks (Easley, O'hara, and Srinivas (1998) and Hu (2014)) as well as on versus off-exchange, which is our innovation. Due to the advantage of leverage, the informed traders would choose options at time $t-1$ if option liquidity is high, i.e., trading at options NBBO is profitable across the proliferation of contracts across strikes and maturities. Options trades alert liquidity providers on both option exchanges and lit stock exchanges about the arrival of new information. The liquidity providers' reaction to the option trade is to widen quoted spread in both markets at time t. This conjecture in our model leads to our first testable hypothesis that expected stock spreads in period t are predicted by options activity. Figure 3 presents initial evidence supporting this hypothesis at trade by trade level. In regressions at the aggregate level, we confirm that stock spreads react to options spreads, volatility and recent changes in option open interest.

In face of higher trading costs on option and stock exchanges, our model predicts that the informed traders continue exploiting their advantage with innovative low cost inside-spread orders in off-exchange trading venues, which also allow stealth trading opportunities due to their lack of pre-trade transparency and delayed post-trade transparency (Ernst, Sokobin, and Spatt (2021)).[4] Informed traders' strategy works only if they can find counterparty liquidity, which they do, because uninformed counterparty traders are also attracted to the lower trading costs in the off-exchange venues. Our second hypothesis tests the relation between expected spreads and off-exchange trading volume and informativeness. The dramatic jump in probability of off-exchange trading in trade by trade analysis after informed options trades, shown in Figure 4 and subsequent regressions at the aggregate level provide strong evidence supporting our second hypothesis. In our third hypothesis, we test if the strength of the second hypothesis's relation between expected spreads and off-exchange trading increases in intense information advantage periods before earnings announcements. Finally, our fourth hypothesis serves as a counterfactual against the null hypothesis of uninformed off-exchange trading by shedding light on the role of

[4]Trades are disclosed publicly in less than 30 microseconds by exchanges and more than 2,500 microseconds by off-exchange trade reporting facilities.

day-to-day variations in off-exchange trading volume and informativeness for predicting stock realized volatility using machine learning methods.

Unlike Comerton-Forde and Putniņš (2015) and Buti et al. (2022) who attribute wider spreads to increased informed trader ratio on exchanges and assume off-exchange trading to be uninformed (Hagströmer and Menkveld (2024)), our paper provides information signal from options as an important additional reason for wider spreads on exchanges and the off-exchange migration of informed trading. The reality is that many off-exchange dark pools are open to anyone including informed traders. The industry views this as a significant problem and therefore some ATS private rooms have emerged to try harder to keep informed traders away from that sub-set of off-exchange interactions.[5]

To capitalize on the valuable insights contained in off-exchange trade reports (Ernst, Sokobin, and Spatt (2021), we use 15-years of intra-day Daily Trades and Quotes (DTAQ) dataset from 2008 to 2022 for a comprehensive sample of all firms trading on the US exchanges. Off-exchange trading is marked in DTAQ dataset with the exchange code of 'D'-FINRA (the Financial Industry Regulatory Authority) in the trade files. We assess the off-exchange information share using both traditional and novel machine learning methods. Following Hasbrouck (1995) and Comerton-Forde and Putniņš (2015), we estimate the off-exchange information share from a vector error correction model (VECM) commonly used in prior literature. To gauge the relative information contribution, we follow Chakravarty, Jain, Upson, and Wood (2012) to define off-exchange information density as the ratio of information share and market share of volume. We also use novel non-parametric machine learning methods such as Extreme Gradient Boosted Tree (XGBoost) for computation of informed trading intensity proposed by Bogousslavsky, Fos, and Muravyev (2024), for model validation, and for assessing the relative importance of explanatory variables using Least Absolute Shrinkage and Selection Operator (LASSO), and SHapley Additive exPlanations (SHAP) values. These methods relax the strict restrictions of traditional models and allow for improved identification of the determinants of informed off-exchange trading.

[5] https://finance.yahoo.com/news/darker-dark-pool-welcome-wall-210011026.html

We empirically examine the cross-market drivers of the market share and informativeness of off-exchange trading to test our four interconnected hypotheses in a two-stage regression consistent with our two-period model. The first stage regression coefficients confirm our models' first prediction that for optionable stocks, the stock-market quoted spreads are positively predicted by the option-market quoted spreads, implied volatility, and option open interest. In the second stage, we test our second hypothesis that off-exchange market share and information share in a given period is higher if the expected NBBO spread on exchanges is higher in that period. We test the our hypothesis both at the daily aggregation of off-exchange market shares to understand the economic significance of the model for investors and regulators and also at the trade-by-trade level to provide full high-frequency granularity into the model's insights. In the daily aggregated regressions, We find that a one percent increase in the predicted quoted spread leads to a 7.13% increase in the off-exchange market share, a 5.83% increase in the information share, and a 0.0438 increase in the information density. At the trade-by-trade level, the results are much stronger as activity in the options market is followed by a dramatic jump in the probability from about 40% to about 60% that the next trade occurs off-exchange.

Our findings at first glance contradict Menkveld, Yueshen, and Zhu (2017)'s findings based on their smaller sample of 117 stocks in October 2010 that informed traders avoid off-exchange trading due to their lower order execution fill rates. To reconcile the differences with prior literature, we conduct sub-sample analyses for three periods, namely 2008-13, 2014-19, and 2020-22. Indeed in the early years (2008-2013), the coefficient for expected spread is negative in the regression with off-exchange information density as the dependent variable. In contrast to this conventional wisdom, we find that in recent years, the off-exchange market share, off-exchange information share, and off-exchange information density all increase with higher expected stock spreads. Furthermore, the results are strongest in the most recent sample from 2020-22. This pattern implies that the immediacy of off-exchange has improved sufficiently in recent years to levels acceptable to informed traders who want to strategically exploit the stealth trading feature of off-exchange dark pools (Ye and Zhu (2020)).

Earning announcements days are times when informed trades exploit their heightened but short-lived information advantage more aggressively. The spike in Figure 7 illustrates that informed traders intensely use low-cost off-exchange venues around earnings announcements. Consistent with our third hypothesis, the impact of wider expected spreads on increasing off-exchange market share and informativeness is stronger during earnings announcement periods, especially for large stocks, further validating the increased price discovery role of off-exchange trading.

Our results about the increase in off-exchange trading's informativeness due to higher on-exchange spreads survive even in the face of strong unconditional on-exchange trading spike headwinds on quad-witching days when futures and various option maturities all expire together. Market shares reshuffle on major derivative settlement days. We are the first to find a recurring pattern of the declining proportion of off-exchange trading because derivative settlement prices are determined on listing exchanges, giving exchanges a competitive lifeline to resist the rise of off-exchange trading. Even with this general market share reshuffling on derivative settlement days, we still infer increased off-exchange informed trading when the predicted quoted spreads on exchanges are higher.

Another analysis that corroborates the use of innovative order types at off-exchange venues by informed traders is our sub-sample analysis based on the amount of price improvement. Sophisticated informed traders are more likely to use their orders at midpoint or with large price improvements. Both of these sub-samples with major price improvements have statistically significant positive coefficients, unlike small price improvement trades which are more likely used by uninformed retail traders.

The importance of options activity highlighted by our first two hypotheses is confirmed in our expanded sample that includes optionable stocks with indicator and interaction variables, and adds non-optionable stocks as a benchmark control group. Market share and information share increase with the predicted stock market spreads and the effect is an order of magnitude higher

for optionable stocks than non-optionable stocks. Additionally, our trade-by-trade probit analysis confirms that the probability of off-exchange trading is higher after an informed option trade, especially large and out of the money option trades, particularly when quoted stock spreads are wider on exchanges.

Finally, we test the volatility implications of the rise of off-exchange market share and informativeness in Hypothesis 4. We are one of the first papers to use off-exchange information to predict future volatility using Machine Learning methods. The literature on volatility forecasting with economic variables has generally not focused on microstructure predictors[6]. Recently, Easley, López de Prado, O'Hara, and Zhang (2021) demonstrate how machine learning can be applied to microstructural research. Rahimikia and Poon (2020) have compared ML models with HAR models for forecasting daily RV by using variables extracted from limit order books and news. By contrast, we are using LASSO with off-exchange information to forecast volatility. In comparison to other black-box machine learning models, LASSO can intuitively and directly show the relative importance of the new proposed variables (e.g. Chinco, Clark-Joseph, and Ye (2019) use LASSO method to predict equity returns). We find that the type of informed traders that choose to trade in off-exchange venues increases the volatility of future prices in the stock markets.

Overall, our paper contributes to several strands of trading literature on the role of off-exchange. On the theoretical side, we build upon the equity exchange framework of Budish, Cramton, and Shim (2015) to include cross-market information from options and off-exchange venues. On the empirical side, the existing literature uses fewer stocks and much shorter sample periods (Menkveld et al. (2017), Brogaard and Pan (2022), Brugler and Comerton-Forde (2022), Buti et al. (2022), Comerton-Forde and Putniņš (2015), Cox (2020)). The insights from our multimarket model and sub-sample analyses of one of the most comprehensive microstructure panel data help us reconcile and extend empirical literature by accounting for the improved immediacy in modern off-exchange trading that allows informed participants to pick up volume in these markets. Our

[6] Xiong, Nichols, and Shen (2015) have applied LSTMs to forecast S&P 500 volatility, with Google domestic trends as predictors. Christensen, Siggaard, and Veliyev (2023) have examined the performance of ML models in forecasting 1-day-ahead RV with firm-specific characteristics and macroeconomic indicators.

paper provides evidence consistent with this conjecture by showing that off-exchange market shares increases with a given stock's informed trading intensity identified by machine learning methods (Bogousslavsky, Fos, and Muravyev (2024)).

Our paper also contribute to the options-equity literature on lead-lag relationship by incorporating signaling effects for liquidity suppliers. Muravyev et al. (2013) show option price quotes do not contain significant information about future stock prices beyond what is already reflected in the stocks. In contrast, Hu (2014) computes option-induced order imbalance and finds it significantly predicts future stock returns in the cross-section on a daily level. Consistent with Hu (2014), we are the first to show that options not only have price leadership but options activity is also a determinant of NBBO quotes and thereby off-exchange market share and informativeness.

In summary, our model and empirical regressions show that when informed traders predict high spread on exchanges, they increase their use of off-exchange trading. We further observe significant increases in off-exchange market shares across all stocks on earnings announcement days, enhancing the price discovery role of off-exchange trading. Increased off-exchange contribution to price discovery may explain why the observed high off-exchanges trading volume market shares exceeding the so-called tipping point of 40% haven't resulted in a collapse or deterioration of the overall price discovery process.[7] Nonetheless, our LASSO analysis shows that increased off-exchange trading predicts stock price volatility in following periods. Thus, our analysis offers a range of regulatory insights with broad and multifaceted implications.

The remainder of the paper is organized as follows. Section 2 provides the background of options and equity markets leading to the development of our hypothesis regarding information content of off-exchange trading in interconnected markets. Section 3 contains the details of the data sources, variable definitions, and empirical methods. Section 4 provides the empirical results and Section 6 concludes.

[7] In testimony before the Senate Banking, Housing and Urban Affairs Committee, Nasdaq OMX Chief Economist,Frank Hatheway, states that based on Nasdaq's empirical analysis "execution quality begins to deteriorate when stocks experience dark trading in excess of 40% of total volume."

2 A Stylized Model and Hypothesis Development

To better understand our empirical evidence highlighting the interconnectedness between the option market considerations and off-exchange stock trading, we build upon the framework Budish, Cramton, and Shim (2015) to develop a two-period multimarket model. In our model, due to the implied leverage in option positions, informed traders find it more profitable to trade options first at normal spreads. Thus, information is first revealed in the options markets. Subsequently, liquidity providers and liquidity demanders in options markets, on stock exchanges, and off-exchange venues observe these trades. Liquidity suppliers earn the spreads from liquidity demanders and face losses against informed traders who trade opportunistically across both markets. Therefore, liquidity suppliers adjust their spreads upwards with changing information environment. Figure 3 shows that this assumption is deeply grounded in reality. Higher spreads, in turn, increase the attractiveness of off-exchange venues where opportunities and mechanism exist to trade at the midpoint or inside the spread with price improvement.

Our two-period model has an informed trader, options market liquidity suppliers, stock market liquidity suppliers, and uninformed traders in both markets. Uninformed traders buy and sell stocks to satisfy their liquidity needs. The informed trader implements trading decisions in the options market first because levered option positions generate higher rewards for her informational edge. Option and stock market liquidity suppliers observe the order imbalance generated by the informed trading in options market throughout period t-1 and adjust their spread wider in each market in the following period to protect themselves against adverse selection. Liquidity suppliers' cross-market informational sensitivity is the novel contribution of our model. [8] Traders and liquidity suppliers take options and stock intrinsic values as exogenous in each period of time. Options prices $O_t = (call_t, put_t)$ and the stock price y_t are related through a Black–Scholes formula, both of which can change over time. For simplicity, we develop the model with only one out-of-money

[8] If the stock does not have an option written on it, the informed trader would use ISO in t-1 and the stock market liquidity supplier would still widen spreads.

call and one out-of-money put as an example.

Period t-1: At time t-1, an informed trader processes private information and trades in the options simultaneously causing a widened quoted spread in the options.

Period t: At time t, given the quoted spreads and implied volatility in the option from period t-1, the stock liquidity supplier makes a decision on the quoted spread for the stock. At the same time, a stock-informed trader makes a decision on trading on or off-exchange or exit.

Informed traders prefer options over stocks in the first period when option spreads are normal as options are more profitable due to leverage (Easley, O'hara, and Srinivas (1998)).[9] However with widened spreads and increased implied volatility, options trades become less profitable in the period t, and off-exchange midpoint trades, off-exchange trades inside the quote, or off-exchange stealth trades may offer better opportunities to informed traders in the period t. All traders are also faced with budget constraints and lot size constraints. Without loss of generality, and primarily for simplicity and tractability, our model makes this assumption of the sequential trading order. In addition, our model does not allow off-exchange tradings of options. This is because all options trading must be executed on exchange. Informed or not, traders are not allowed to trade options off the exchange. Finally, our model assumes there are two liquidity suppliers. In reality, the two liquidity suppliers may collude or one liquidity supplier may provide liquidity for both options and stocks (on-exchange and off-exchange).

Uninformed Traders: Assume that the option uninformed trader arrives with an arrival rate λ^{O}_{Invest} for each period with double symmetric demand — half of them are buy orders and the other half are sell orders, and half of these buy orders are calls and the other half buy orders are puts at the midpoint price $O = (call, put)$. Given the options liquidity supplier's ask price $A^{O} = (A_{call}, A_{put})$ and bid price $B^{O} = (B_{call}, B_{put})$, the dollar transaction cost for uninformed option traders is the

[9] ISO are used instead of options for non-optionable stocks with similar implication for period t.

sum of the cost of buying and selling options (effective spread),

$$C^O_{buy} = \frac{\lambda^O_{Invest}}{4}[(A_{call} - call) + (A_{put} - put)], \quad (1)$$

$$C^O_{sell} = \frac{\lambda^O_{Invest}}{4}[(call - B_{call}) + (put - B_{put})]. \quad (2)$$

The uninformed options traders choose the child size order q^O and arrival rate p^O_q to minimize the total transaction cost $C^O_{buy} + C^O_{sell}$ each period, while $\sum_{q^O=1}^{\bar{q^O}} q \times p^O_q = \lambda^O_{Invest}$. In equilibrium, $p^O_q = 0$ for any child size order larger than one unit. The options liquidity supplier would update the ask and bid price after each trade.

Similarly, stock uninformed traders arrive at a rate of λ^S_{Invest} for each period again with symmetric demand with half buy orders and half sell orders. Given the stock liquidity supplier's ask price A^S and bid price B^S, the dollar transaction cost for uninformed stock traders is the sum of the cost of buying and selling stock (effective spread),

$$C^S_{buy} = \frac{\lambda^S_{Invest}}{2}(A^S - y_t), \quad (3)$$

$$C^S_{sell} = \frac{\lambda^S_{Invest}}{2}(y_t - B^S). \quad (4)$$

The uninformed stock traders choose the child size order q^S and arrival rate p^S_q to minimize the total transaction cost $C^S_{buy} + C^S_{sell}$ each period, while $\sum_{q^S=1}^{\bar{q^S}} q \times p^S_q = \lambda^S_{Invest}$. In equilibrium, $p^S_q = 0$ for any child size order larger than one unit. The stock liquidity supplier would update the ask and bid price after each trade.

Informed Traders: Informed traders know y_t at time $t-1$ and they profit from adversely selecting the options liquidity supplier at time $t-1$. Given the bid-ask spread $s^O_{t-1} = A^O - B^O$ in the options market, informed traders would sweep all outstanding bids (asks) for the OTM put and outstanding asks (bids) for OTM calls if y_t is significantly above (below) the y_{t-1} overcoming the transaction cost $\frac{s^O_{t-1}}{2}$.

Suppose that right before the stock market opens at t^-, there is still a large enough gap between y_{t^-} and y_t, potentially overcomes the transaction cost $\frac{s_t}{2}$, where $s_t = A^S - B^S$. For stocks with a rather liquid options market, this second-period arbitrage opportunity is thin. This is because first, the option implied price of the underlying should be very close to the true y_t at the time t^-. Second, the liquidity provider in the stock market updates its quotes in the stock market very quickly. Therefore, if $y_{t^-} - \frac{s_t}{2} < y_t < y_{t^-} + \frac{s_t}{2}$, the informed traders would choose to trade off-exchange at time t.

Equilibrium Bid-Ask Spread: Inspired by Budish, Cramton, and Shim (2015), the bid-ask spread of the stock s_t leaves liquidity provider with zero profits,

$$\lambda^S_{Invest,t} \times \frac{s_t}{2} = \lambda^S_{Jump,t}(Z_{t^-}) \times [E(J|Z_{t^-}) - \frac{s_t}{2}], \tag{5}$$

where Z_{t^-} is the information set of the stock liquidity supplier at time t, and J is the jump size in the price $J = y_{t^-} \times (e^x - 1)$, assuming the dynamic continuous-time representation for the instantaneous arithmetic return on y.

$$\frac{dy_t}{y_{t^-}} = \int_{\Re} (e^x - 1)\mu(dt, dx), \tag{6}$$

where $\mu(dt, dx)$ is a counting measure for the jumps in y with compensator $dt \bigotimes \nu_t(dx)$, $\nu_t(dx) = \lambda^S_{jump,t} \times f(x)$, and $f(x)$ is the probability density function of the jump size x.

At time t, the stock liquidity supplier can observe revealed information on stocks and trading activity in options, such as stock volatility σ_{t-1}, implied volatility $\sigma^{Imp}_{t-1,t}$, and the summation of net uninformed open interest $\Delta OI_{Invest,t-1}$ and net informed open interest $\Delta OI_{Jump,t-1}$. Without loss of generality, we define the information set $Z_{t^-} = (\sigma_{t-1}, \sigma^{Imp}_{t-1,t}, \Delta OI_{Invest,t-1} + \Delta OI_{Jump,t-1})$. Rearranging the above equation, we have the bid-ask spread s_t as the following,

$$s_t = \frac{2 \times E(J|Z_{t^-})}{1 + \frac{\lambda^S_{Invest,t}}{\lambda^S_{Jump,t}(Z_{t^-})}}. \tag{7}$$

At time $t-1$, suppose the option implied volatility is high coupled with an abnormally large open interest occurs in the options market, then the liquidity supplier's ex-ante expectation of the arrival rate $\lambda^{S}_{Jump,t}(Z_{t^-})$ and expected jump size $E(J|Z_{t^-})$ both increase. To gain protection and reduce the chance of being adversely selected, the stock market marker thus increases the bid-ask spread s_t.

Hypothesis 1: For optionable stocks, the stock quoted spreads should be positively predicted by the options quoted spreads, option implied volatility, and change in open interest.

Higher quoted spreads in options and stock exchanges force informed traders to consider alternative trading venues, such as off-exchange. Based on the supporting anecdotal evidence in Figure 4, we formally test the following hypothesis for all stocks:

Hypothesis 2: The off-exchange market share and the information share increases with stock spreads, particularly for optionable stocks. Informed traders trade in the options market at time $t-1$ first. Consequently all spreads widen; stealth and cost sensitive trading shift to off-exchange at time t.

Stealth trading motivation also make off-exchange trading attractive for informed traders (Ye and Zhu (2020) and Topbas and Ye (2024)) as an alternative to order splitting (Chakravarty (2001)) due to the less transparent off-exchange trading environment. Trading on exchanges at bid or offer prices quickly reveal the informed traders intentions to the liquidity suppliers. In contrast, off exchange dark pool trading at midpoint crossing networks empower informed traders to minimize or eliminate price impact allowing them maintain their information edge for a longer period and execute larger quantities of profitable trades at favorable prices.

3 Data and Methodology

We use intraday trade-level data for our analysis, covering the period from January 1, 2008 to December 31, 2022. Our primary data source is the Daily Trades and Quotes (DTAQ) dataset, which is accessible through the WRDS platform. To ensure data consistency, we focus solely on

common stocks listed on the NYSE, NYSE MKT, and NASDAQ, with CRSP share codes 10 or 11 and exchange codes 1, 2, or 3. This results in a universe of 8,184 firms with DTAQ data. We investigate the trading behavior and price discovery in both on- and off-exchange markets during regular trading hours. To identify off-exchange trading, we rely on the DTAQ dataset, which marks such trades with an exchange code of "D", signifying FINRA (the Financial Industry Regulatory Authority), which is responsible for compiling and reporting off-exchange trading to Securities Information Processor (SIP). Moreover, we follow the method proposed by Holden et al. (2023) to adjust exchange-to-SIP latency for all trades from August 2015 onward.

We conduct our analysis by combining the DTAQ dataset with additional data sources. We identify quarterly earnings announcements using the announcement date and times recorded in the Thomson Reuters I/B/E/S database. Following, Patton and Verardo (2012), DeHaan, Shevlin, and Thornock (2015) and Chan and Marsh (2022), we assign a new event day for the announcements occurring after market close so that "day 0" in our event window is the day in which investors can react to the earnings announcements.

Furthermore, we obtain daily option prices and volume from OptionMetrics. We identify a stock as optionable if it has at least one record in OptionMetrics during our sample period, and as non-optionable otherwise. Of the 8,184 distinct firms, 5,591 (68.3%) are optionable and 2,593 (31.7%) are non-optionable. We also obtain the intraday volatility from Millisecond Intraday Indicators Database by WRDS as control variables.

For initial deep-dive evidence, we focus on the Amazon stock for full visual granularity at the trade-by-trade level. We obtain additional 2.87TB of put and call option trade and quote OPRA data for the entire option chain of all maturities across all strike prices for all trading days in 2020 from ThetaData. For the aggregate regression analyses for all stocks for all days in the sample period, we construct a daily time series of three firm-level off-exchange measures to investigate the dynamics of off-exchange trading activities.

1. $MS_{i,t}^{\text{off}}$ — Market Share

We determine the off-exchange market share by computing the ratio of the off-exchange share volume to the total share volume across all trading venues. Our primary focus is on the trading activity during regular trading hours. Therefore, share volume is calculated based on trades with a timestamp on DTAQ between 9:30 AM and 4:00 PM. There are 12.8 million firm-day off-exchange market share observations in our sample.

2. $IS^{\text{off}}_{i,t}$ — Information Share

Following Hasbrouck (1995) and Comerton-Forde and Putniņš (2015), we estimate the following vector error correction model (VECM) for each firm-day using one-second intervals, t:

$$\begin{aligned}
\Delta p_{\text{off},t} &= \alpha_1 (p_{\text{off},t-1} - p_{\text{on},t-1}) + \sum_{i=1}^{60} \gamma_i \Delta p_{\text{off},t-i} + \sum_{j=1}^{60} \delta_j \Delta p_{\text{on},t-j} + \varepsilon_{\text{off},t} \\
\Delta p_{\text{on},t} &= \alpha_2 (p_{\text{off},t-1} - p_{\text{on},t-1}) + \sum_{i=1}^{60} \phi_i \Delta p_{\text{off},t-i} + \sum_{j=1}^{60} \varphi_j \Delta p_{\text{on},t-j} + \varepsilon_{\text{on},t}
\end{aligned} \tag{8}$$

where $p_{\text{off},t}$, and $p_{\text{on},t}$ are the log of last available traded price in off-exchange and on-exchange, respectively. We compute $IS^{\text{off}}_{i,t}$ and $IS^{\text{on}}_{i,t}$ for firm i on day t from the error correction parameters and variance-covariance matrix of the error terms, following Baillie, Booth, Tse, and Zabotina (2002). In cases where there is no or only a limited number of on-exchange or off-exchange trades, the estimates may not converge. In such instances, we set the information share as blank. Ultimately, we have 8.3 million firm-day converged off-exchange information shares in our sample.

One implication of increased information share of off-exchange trades is that the liquidity supplier's realized spread from those trades should decline. Following Conrad and Wahal (2020) we compute the probability of realized spread being positive for each off-exchange trade. Realized spread ($rs_{i,t,\tau}$) is defined as the trade price ($p_{i,t}$) minus the NBBO midpoint

after the trade times ($m_{i,t+\tau}$) as follows:

$$rs_{i,t,\tau} = \frac{q_{i,t}(p_{i,t} - m_{i,t+\tau})}{m_{i,t}}, \tag{9}$$

where $q_{i,t}$ is equal to +1 for buyer-initiated trades and -1 for seller-initiated trades. The probability of realized spread being positive is computed as the number of off-exchange trades with positive realized spread divided by the total number of off-exchange trades on that day. This measure serves as a confirmatory measure of the impact of rising off-exchange information share. An additional benefit of this measure is that it can be computed for each off-exchange trade whether or not the Hasbrouck VECM measure converges for that day.

3. $ID^{\text{off}}_{i,t}$ — Information Density

   Following Chakravarty, Jain, Upson, and Wood (2012), we define the off-exchange information density as

$$ID^{\text{off}}_{i,t} = \frac{IS^{\text{off}}_{i,t}}{MS^{\text{off}}_{i,t}} \tag{10}$$

   where $IS^{\text{off}}_{i,t}$ is the off-exchange information share for firm i on day t and $MS^{\text{off}}_{i,t}$ is the off-exchange market share for the corresponding firm on a given day. Similarly, we can construct the on-exchange information density. We compute information density for 8.3 million firm-days in our sample.

4. $ITI_{i,t}$ — Machine Learning Based Informed Trading Intensity

   In addition to using the traditional information share measures described above, we also take advantage of recent advancements in identification of informed trading intensity (ITI) for which Bogousslavsky, Fos, and Muravyev (2024) train a machine learning method on a class of informed trades, namely 13D filers on SEC EDGAR, who are expected to have a higher signal-to-noise ratio according to Collin-Dufresne and Fos (2015). The scikit-learn package in Python helps implement Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest and provides an XGBoost interface for the Gradient Boosted

Trees (GBT) algorithm recommended by Bogousslavsky, Fos, and Muravyev (2024). In addition to the 41 predictors used by them, our goal is to utilize additional trading venue-specific microstructure variables to assess the informativeness of off-exchange versus on-exchange trades.

In 2007, the SEC approved a rule change that allowed investors to bypass exchange tick sizes to gain sub-penny price improvements from off-exchange trading venues.[10] This rule change stimulated competition among trading venues and contributed to the growth of off-exchange market. Figure 1 Panel A displays the time series of daily off-exchange market shares from 2008 to 2022. Trading in off-exchange venues has become increasingly popular in the last 15 years, with the increase in its market shares from 23.59% in January 2008 to 41.62% in December 2022.[11] The rising trading volume market share of off-exchange venues is particularly notable despite the headwinds and regular downward spikes on derivative settlement days, which is another novel result that we discuss in greater detail in the empirical results section. Due to the Covid-19 pandemic, the advent of commission-free trading platforms such as Robinhood, and the MEME stock frenzy, there has been an even larger increase since 2020 in trading by retail investors', whose orders are preferred so much by the off-exchange internalizers that they share some of their spread in form of payment for order flow (PFOF) to the brokers. In addition to full sample analysis, we also provide sub-sample analyses for the three distinct periods of 2008-13, 2014-19, and 2020-22 which have different levels of market share visible in Panel A.

Because off-exchange trades are priced based on national best bid and offer (NBBO) from exchanges which are important contributors to price discovery and market quality (Brogaard, Ringgenberg, and Roesch (2023)), the information share of off-exchange trades and their limited role in price discovery is an important economic question. Figure 1 Panel B depicts the time series of off-exchange information share. Contrary to traditional wisdom, we find that informed investors

[10] SEC Rule SR-NASD-2005-146 expanded the scope of IM-2110-2 allowing off-exchange venues to provide sub-penny price improvement for exchange listed securities.

[11] Off-exchange market share in 2023 increased to 43.97%, surpassing 2021's record of 43.64% according to Rosenblatt Securities (2023). See also the summary in CBOE U.S. Equities Market Volume Summary.

share of off-exchange trades has risen from an average of 11.57% in the early sample (2008-2010) to 18.67% in the later sample (2020-2022) as depicted by the blue line. The confirmatory implication of this rise in information share also holds true in the declining proportion of trades for which liquidity suppliers are able to earn positive realized spreads depicted by the gray line.

Table 1 reports the average daily trading activities on and off-exchange. Panel A shows the dollar and share volume off-exchange and on-exchange (NYSE, NASDAQ, and other lit-exchanges). Panel B shows the information share and information density off-exchange and on-exchange. First, we see that trading in off-exchange markets constitutes a significant portion of the total trading volume for all stocks averaging 57.44 billion dollars and 1.68 billion shares, accounting for 37.12% of the total share volume over our 15-year sample period. As most of the dollar volume and share volume in concentrated in optionable stocks, we provide more detailed descriptive statistics for optionable stocks from options market, off-exchange venues, and on-exchange in the next table. Overall, off-exchange venues account for nearly one-third of the trading volume for S&P 500 stocks and optionable stocks. For firms that are not included in the S&P 500 index and those without optionability, off-exchange market shares are even higher, reaching over 50%. In Panel B, off-exchange information share in the full sample period is 15.47%, which is much below off-exchange trading volume's market share of 37.12% in the previous Panel. As a result off-exchange information density is below 1. Correspondingly, the on-exchange information shares for the full sample period is 84.53% yielding an on-exchange information density higher than 1. This is consistent with the traditional wisdom that informed traders are clustered on exchanges. However, these static averages hide the fact that off-exchange trading volume shares and information shares have both risen substantially in recent times as we discuss later in the paper. It is also noticeable that for non-optionable stocks, the off-exchange information share is 24.46% than the overall average. The differences in information share and information density motivate us to differentiate optionable stocks from non-optionable stocks, and large firms from small firms.

Table A2 shows the descriptive statistics for optionable stocks and their options activity separately, because optionable stocks represent over 98% of the trading volumes on and off-

exchange. We report all the statistics by firm size because we also learned from the previous table that off-exchange activity for big firms is different from that for small firms. Although optionable stocks cover a large range of firm sizes from an average 0.12 billion to an average 17.84 billion in market capitalization, the average firm size of optionable stocks is larger than that of non-optionable stocks. Large firms attract the majority of the trading volumes of 128.26 million dollars (2.8 million shares), while the medium and small firms combined have trading volumes only 9.57 million dollars (1.17 million shares) each day. We see similar trading concentration by firm size both on and off-exchanges. Average transaction costs on exchanges are 0.28% ranging from 0.12% for large firms and 0.92% for small firms. Options, on the other hand, are almost 200 times more expensive to trade than stocks, and 88 times more volatile. On average, options traders pay 49.77% in transaction costs. However, options have leverage and are also very liquid for large optionable firms. For example, large firms' options have 35 thousand open interest and 5 thousand trading volume, which is about 76 million dollars in notional volumes at the applicable strike price. These options trading volumes are heavily right-skewed with a median value significantly lower than the mean, indicating an inactive options market on a normal day, but massive tradings on possible information release days. Due to the high trading costs, market participants do not trade options if they do not have a strong desire to hedge or to speculate the direction of the stock price. Lastly, during the earnings announcement days, trading volumes of stocks and options both skyrocket to a much higher level, regardless of the firm size. The total stock (options) trading is 5.51 million shares ($10,846$ contracts) per day, which is three times more than a normal trading day. The quoted spreads of stock and options are both higher on earnings announcement days. During the Quad Witching Days, trading volumes of stocks and options both mildly increase but are associated with narrower spreads.

4 Empirical Results

We begin our analysis by examining the relationship between off-exchange activity and machine learning-based measures of informed trading intensity. Evidence from this analysis serves as our initial evidence that informed traders utilize off-exchange venues and motivates our investigation of mechanisms and specific determinants of off-exchange activity. We then proceed to test our four hypotheses to explore the role of lower-cost and stealth trading motivations in driving daily variations in off-exchange trading.

4.1 Granular Trade-Level Analysis

4.1.1 Initial Test of the Model

Figure 3 shows the changes in dollar quoted spread in the options market in the left panel and for stock market in the right panel, following a option trade execution. We plot the figure for 25.02 million option trades in the Amazon stock for all trading days in 2020 for visual clarity and computational ease at the trade-by-trade level, but later follow up with regressions after aggregating the information at the daily level for all stocks for all years in the sample period. For the figure itself, we collect the option quoted spread at 100 millisecond intervals from five seconds before to five seconds after each option trade. The figures show a remarkable increase in the average dollar quoted spread surrounding the option trade both for option contracts in the left panel and for the underlying stock in the right panel. This jump in spreads provides initial evidence supporting Hypothesis 1 that liquidity providers in both markets widen quoted spread after they infer presence of informed participants from option trades.

For our initial trade-level test of Hypothesis 1, we regress the stock spreads at the time of each trade on the previous minutes' options volume and log price with control of day fixed effects. Table 2 Panel A reports the predictive results from above regression at 1, 5, 15, and 30 minute

aggregations for dollar quoted spread on the left and percent quoted spreads on the right. All coefficients for option volume in the first row are positive and statistically significant. Thus, options activity predict an increase in stock spreads for both measures at all frequencies, consistent with Hypothesis 1.

We also use the trade level data to conduct the initial test of our model's Hypothesis 2 in the 2nd stage regressions in Panel B with off-exchange market share as dependent variable and the predicted stock spread from stage 1 as the key explanatory variable. All coefficients in the first row are positive and statistically significant for 1, 5, 15, and 30 minute aggregations in these regressions as well. Thus, an increase in the predicted quoted spread leads to an increase in the off-exchange market share, consistent with Hypothesis 2. In the next section we validate our model and findings with machine learning methods and later, we test both hypotheses with the full universe of 8,184 stocks and additional cross-sectional control variables in the 2 stage regression equations described in the section 4.2, where we also discuss the economic significance of our findings.

4.1.2 Machine Learning Validation

To validate our model and findings, we want to conduct an independent test of the impact of option trading on trading venue choice of following trades without imposing the model's assumptions or parametric restrictions of the 2 stage OLS regressions. Figure 4 presents the direct visual evidence of the impact of typically more informed option trades on the likelihood of off-exchange trading at the trade-by-trade level to provide full high-frequency granularity with one stock for our cross-market model's composite insights. The figure presents the unconditional intraday probability of a trade occuring off-exchange with blue line and the conditional intraday probability of a trade occuring off-exchange following the execution of an options trade with orange line. The unconditional probability is calculated as the ratio of off-exchange trades to total trades aggregated from trade-by-trade data for each day in shown in the plot, while the conditional probability is determined based on the first stock trade occurring off-exchange after an options trade.Consistent

with our model's composite predictions, activity in the options market is regularly followed by a dramatic jump in the probability from about 40% to about 60% that the next stock trade occurs off-exchange. The probability of on-exchange trade (not directly shown in the figure) is 1-P(off-exchange) and it drops to less than half after an options trade, highlighting the dominance of off-exchange trading. Following this initial visual evidence, we prepare probit models which can allow the use machine learning techniques to validate the choice of variables in our theory model.

In Table 3, we start to estimate probit regressions of the trade-by-trade probability of the trade occurring off-exchange instead of occurring on an exchange. The equation for the pooled probit regression including the control variables is as follows:

$$\begin{aligned}\text{Prob(Off-exchange)} = &\alpha + \beta_1 I_{\text{After option}} + \beta_2 I_{\text{Expiration Date}} + \beta_3 \text{Last Quoted Spread} + \\ &\text{Last Option Trade Information} + \text{Last Stock Trade Information} + \\ &\text{Other Controls} + \varepsilon\end{aligned} \tag{11}$$

The dependent variable is an indicator variable equal to 1 if the trading venue is off-exchange and 0 if it is on an exchange. Last option trade information includes option trade price, option trade size, Call/Put indicator, time to maturity, option moneyness, option quoted spread, option's NBO and NBB depth. Last stock trade information includes stock trade price, stock trade size, off-exchange trade indicator. Other controls includes last 1 second and 5 second stock quoted spread change, minute indicator, stock NBO and NBB depth. The key explanatory variable is an indicator variable, After Option, which takes a value of 1 when the stock trade is preceded by an options trade and 0 otherwise. With day fixed effects akin to separate regression for each day, the positive and statistically significant (at the 1 percent level) average coefficient of 0.5938 on after options trade variables strengthens the evidence supporting H2 that off-exchange market share and information share rises after options trading. Moreover, to understand the economic significance of this result, we show that a one standard deviation or 0.91 dollar spread increase leads to 0.147×0.91=13.37% increase in the probability of trading off-exchange.

The coefficient estimate of 0.5895 in the pooled Probit regression with nearly 46 million observations is very similar to the values plotted in the figure and coefficient without the control variables. Thus, the trade-by-trade level analysis with the full sample of all AMZN trades reinforces the notion that off-exchange market share and information share rises after options.

In the internet appendix, we provide a robustness test by focusing only on sub-sample of off-exchange stock trades that occur right after an option trade and confirm the role of options and spreads in trading venue choice.

Based on our model, investors' decisions to route trades to off-exchange are primarily informed by the most recent stock trade, option trade, and prevailing market quote information. This suggests that a machine learning model can be trained on this information set to predict whether the next stock trade will be executed on off-exchange. Such a model is expected to achieve high out-of-sample accuracy, with option and quote information playing a central role in its predictive power. We compile all stock trades from three trading days before to three trading days after each earnings announcement, yielding a 28-day sample comprising 5.36 million observations. Of these, 2.5 million trades are randomly sampled to train the XGBoost model, and the remaining 3.36 million trades are used for validation.

To evaluate the marginal importance of each feature, we compute the SHAP value for every variable across all out-of-sample observations following Bali, Beckmeyer, Moerke, and Weigert (2023). The SHAP values provide an additive decomposition of the model's prediction, quantifying each variable's marginal contribution, positive or negative, to the predicted outcome. The prediction for each out-of-sample observation k can be expressed in log-odds form as

$$logit(p(k)) = \beta_0 + \sum_{i=1}^{19} \text{SHAP}_i^k, \tag{12}$$

where $p(k)$ is the predicted probability, β_0 is the baseline log-odds, and SHAP_i^k denotes the SHAP value of variable j for observation k.

Panel B of Table 3 reports the summary statistics of the absolute SHAP values based on 3.53 million out-of-sample observations. From equation (12), a higher absolute SHAP value for a given variable indicates a larger marginal contribution to the predicted outcome. Accordingly, we rank the variables by the mean of their absolute SHAP values. The four most influential features in our XGBoost model are: the off-exchange indicator for the last stock trade, the last quoted spread, the one-second change in the quoted spread, and the post-option dummy variable. These four variables exhibit mean absolute SHAP values ranging from 0.17 to 1.20, markedly higher than the fifth-ranked variable, trade size, which has a mean of only 0.10. Moreover, these key variables display similar means and medians, suggesting limited skewness in their distributions. The variable importance derived from this unsupervised approach aligns closely with our theoretical predictions.

To further validate the model-implied relationships, we plot two key variables supported by both the machine learning results and our theoretical framework: the post-option trade dummy and the last stock quoted spread. Figure 5 shows that all out-of-sample observations with a post-option trade dummy exhibit significantly positive SHAP values, while those without exhibit negative SHAP values. According to equation (12), a positive SHAP value increases the probability of off-exchange execution, consistent with the theoretical prediction. Similarly, the scatter plot for the last stock quoted spread indicates that observations with narrow spreads generally have negative SHAP values, implying a lower probability of off-exchange execution. More importantly, SHAP values increase monotonically with the quoted spread, suggesting that a wider spread raises the likelihood of the next trade being routed off-exchange. Together, these patterns strongly support the conclusion that investors' routing decisions are primarily driven by option market signals and stock quote conditions, in line with our theoretical model.

Having validated the importance of variables included in our theoretical prediction with machine learning techniques, we will formally test the model predictions in 2-stage regressions using all stocks in the next section.

4.2 Predicting Stock Spreads with Options Activity for All Stocks

In this sub-section, we test whether our **Hypothesis 1** holds true for the broader universe of all optionable stocks in the TAQ dataset based on the trade-off between higher on-exchange execution probability versus lower-cost off-exchange stealth execution. In the first stage, we regress the stock spreads (time-weighted average percentage quoted spread) on the previous day's options quoted spread (open interest-weighted average percentage quoted spread), at-the-money (ATM) implied volatility and change in open interest with control of time trend and firm fixed effect.[12]

$$Spread_{i,t}^{Stock} = \alpha_i + \beta X_{t-1}^{options} + \beta Trend_t + Fixed_i + \varepsilon_{i,t}. \tag{13}$$

Table 4 reports the predictive results from above regression.[13] The specification (1) to (3) consider three measures of $X_{t-1}^{options}$ from options market separately: $\text{Spread}_{i,t-1}^{Option}$, $\text{Volatility}_{i,t-1}^{ATM}$ and $\Delta\text{Open Interest}_{t-1}$. In column (1), we observe that an increase of 1% in the option quoted spreads induce a 0.08 basis point increase in the stock quoted spread at the 1% significance level. In column (2), an increase of 1% in option implied volatility results in a 0.24 basis point increase in the stock quoted spread. Unlike columns 1 and 2 where the independent variable is expressed in fractions, column 3 is expressed in millions. In column (3), a million contract increase in option open interest causes a 3 basis point increase in the stock quoted spread. This finding is consistent with our H1. On the previous day t-1, the stock liquidity supplier observes the options quoted spread, changes in option open interest, and implied volatility (an ex-ante measure of the volatility). If the options quoted spread, change in open interest, or the implied volatility is high, signaling the adverse selection in the options market, the stock liquidity supplier would widen the stock quoted spread on day t.

[12] The percentage quoted spread (stock or options) is the ratio of the quoted spread in dollars to the midpoint quoted price.

[13] All variables in these regressions with the full universe of stocks are first aggregated at daily level to address additional computational constraints. To bridge these slight differences between granular intra-day analysis and aggregated analysis for all stocks and we did confirm that the inferences for the AMZN stock are the same at both granular and daily aggregation levels as noted below in the footnote for Panel B of Table 5.

For the full regression in the last column, in addition to the options market variables from our model, we follow Stoll (2000) and add control variables including volatility, log dollar trading volume, log market capitalization, log price, and log number of trades, to predict the current day's stock spreads. We also include the previous day's spreads to account for the serial correlation of the spreads:

$$\begin{aligned} Spread_{i,t}^{Stock} =& \alpha_i + \beta_1 \text{Spread}_{i,t-1}^{Option} + \beta_2 \text{Volatility}_{i,t-1}^{ATM} + \beta_3 \Delta\text{Open Interest}_{t-1} \\ &+ \beta_4 \text{Spread}_{i,t-1}^{Stock} + \beta_5 \text{Volatility}_{i,t-1}^{Intraday} + \beta_6 \text{Volume}_{t-1}^{Stock} + \text{Trend}_t + \text{Controls}_t^* + \varepsilon_{i,t}. \end{aligned} \tag{14}$$

Consistent with the literature, we find that the stock quoted spread is positively serially correlated and positively predicted by the past stock volatility. We find that the stock and option spreads, option implied volatility and change in open interest continue to have strong predicting power for stock spreads. The coefficient on stock volatility control variable becomes insignificant in this regression. Altogether, we can explain 23.49% of the variation in stock spreads in the predictive regression.

4.3 Predictable Off-exchange Activities

4.3.1 OLS regressions

To test **Hypothesis 2** in the 2nd stage for the second period, we run regressions for off-exchange market share, information share, and information density on the predicted spreads obtained from the first stage.

$$\text{Off-exchange}_{i,t} = \alpha_i + \beta_1 \widehat{\text{Spread}}_{i,t}^{Stock} + \beta_2 \text{Trend}_t + \text{Controls}_t + \varepsilon_{i,t}. \tag{15}$$

Table 5 Panel A reports the results of the second stage using the universe of all optionable stocks in TAQ.[14] The first column for each off-exchange dependent variable shows that the coefficient for predicted spread by itself is positive and statistically significant. The second column for each off-exchange dependent variable shows the full regression with time trend and control variables. We show that a one percent increase in the predicted quoted spread leads to 7.13% increase in the off-exchange market share, 5.83% increase in the off-exchange information share, and 0.0438 increase in the off-exchange information density.[15] This suggests that a higher expected trading cost on exchange motivates traders in general, and more so the informed traders to trade off-exchange.

The analysis in Panel A drops the observations with negative predicted stock spreads as that would imply crossed markets discussed in Shkilko, Van Ness, and Van Ness (2008) and are typically eliminated from average spread calculations suggested by Holden and Jacobsen (2014). However, our results are robust to this choice as shown in Panel B, where we winsorize the spreads to the minimum value of 0 to keep all of the available observations in the sample. The predicted spread coefficients in Panel B remain positive and significant and have higher magnitudes than Panel A in all 3 regressions. We conservatively use Panel A sample for further analysis.

To understand the economic significance of this result for off-exchange market share, recall from Table 1 that option spreads are very large and volatile. Option quoted spreads from that table are almost 200 times larger (almost 100 times more volatile) than the stock quoted spreads. One standard deviation increase in the option quoted spreads (0.3275 in third column of Table 1 Panel C percent quoted spread row for option market) results in 0.98 basis points in stock spreads (multiplying by the Table 4 Panel A regression coefficient (=0.3275×0.0003) , which in turn leads to 6.99 basis points increase in off exchange market share (=0.98×7.13 coefficient in Table 4 Panel A). This increase is $48,751 dollar volume (6.99 bps $\times$$69.77 million from Table 1 Panel

[14]To bridge the gap between granular trade level analysis and aggregated analysis for all stocks, we confirm that the nferences are the same for all horizons 1,5,15,30 minutes as well as daily aggregation for the AMZN stock. Specifically, the predicted spread has positive and statistically significant coefficients of 207.6 at the 30 mins aggregation level in Panel B of Table **??** reported previously and 60.72 in the internet appendix at the daily aggregation of AMZN stock only to predict the off-exchange market share and information share

[15]Spread, market share and information share are in percent but information density is a ratio.

C) per stock, or multiplied by 2,312 optionable stocks in sample is 113 million dollars increase in total off-exchange volume per day on average.

The signs and significance of coefficients for both Hypotheses 1 and 2 for 8184 stocks with daily aggregation are consistent with the granular trade level results presented earlier, which reinforces the evidence supporting the model's predictions that the off-exchange market share and the information share increases with predicted stock spreads.

In Panel B, we divide the sample into 3 sub-periods. The first two periods align with the sample periods used in prior literature and the most recent third period helps us improve our understanding of off-exchange trading that has evolved rapidly after 2019.[16] Higher expected stock spreads consistently increase off-exchange market share and information share in all three periods. However, the expected spread coefficient in information density regression in the oldest period (the first sub-period) is not positive. This means that in 2008-2013 informed traders pursued off-exchange venues less aggressively and were willing to trade at higher costs exchanges consistent with notion of low off-exchange execution probability Menkveld et al. (2017). But in more recent periods the willingness of informed traders to use off-exchange venues to avoid higher expected spread is evident and consistent with notion of increased information content in Ernst et al. (2021) in 2019-20. The information content of off-exchange trading and the desire to avoid higher expected spread costs has shot up significantly in the most recent periods of 2020-22, which we are the first ones to analyze in the context of off-exchange informativeness.

4.3.2 LASSO regressions

To validate our model and the key result about the jump in off-exchange trading market share and information share, following options activity, we employ LASSO unsupervised machine learning technique which allows for adaptive identification of the most impactful explanatory variables from the full list of candidate variables free of restrictive full sample linear assumptions (Chinco, Clark-

[16] Jain et al. (2020) highlight increased routing of order flow away from exchanges to off-exchange liquidity providers.

Joseph, and Ye (2019)). Table 6 shows the results for off-exchange market share in the left Panel and the Hasbrouck off-exchange information share in the right panel. For each day t, the model is trained on the preceding 750 trading days to validate the model for the following 20 trading days. The sample period covers January 1, 2008, to December 31, 2022 which yields 150 training windows. Our training procedure uses a 5-fold cross-validation method which is a common practice in statistics to select the penalty coefficient in LASSO.[17] We report the summary statistics of the 150 coefficients and the percent of times that the specific predictor variable is selected by LASSO. Within each panel we show the full sample result with 150 training windows and the most recent post-covid and post zero commission 2020-2022 sample period with 36 training windows, to validate our model after these major structural changes. We interpret the machine learning coefficients in Table 6 to indicate their role in directly predicting off-exchange market share and information share, unlike the two-stage regressions in the previous sub-section where they are used to predict the spread changes in the first stage and only indirectly affect the off-exchange activity in the 2nd stage.

Machine learning methods continue to provide strong evidence supporting our hypotheses. The median coefficients on both option and stock spreads are positive consistent with H1 in the full sample analysis as well as the most recent period. The option spread is always positive and is also selected as a predictor variable for higher off-exchange market share as well as information share in 100% of the 150 training models. Previously we saw that implied option volatility increases spread. Although, it did not directly increase off-exchange trading in early years, its role in increasing off-exchange market share and off-exchange information share is consistent with our hypothesis and the rise in off-exchange trading in the most recent post-covid and post-zero-commission era when all market participants have improved access to option trading.

Stock spread is also selected as a predictor variable for higher off-exchange market share

[17] 5-fold cross-validation method refers to following five steps: (1) partition the training set into 5 groups. (2) take out one group as a in-sample validation set. (3) train the model on the remaining data. (4) make predictions on the validation set. This procedure is repeated 5 times (once for each in-sample training set) for a pre-determined penalty coefficient λ. Finally, we select the penalty coefficient with minimum 5-group total mean square error (MSE).

in 97.33% of the 150 models for off-exchange market share and 99.33% of for off-exchange information share in the full sample. In the most recent period, stock spread is selected 100% of the time to explain both off-exchange market share and information share. Signs for liquidity prediction variables proposed by Stoll (2000) price, stock volatility, and number of trades are consistent in traditional model and also directly predict higher off-exchange trading and information share in the machine learning methods.

4.4 Additional Informed Trading Evidence and Expanded Sample

4.4.1 Order Flow by Trade Type

In Figure and Appendix Figure A1 and Figure A2, we provide additional details about changes in off-exchange and exchange-wise market shares for NYSE, Nasdaq (Tape A, B, C), and other exchanges. Figure , Panel A shows that the market share of ISO trades on exchanges has declined whereas the non-ISO share is stable. ISO trades are typically used by informed institutions. Together these trends indicate that off-exchange trading venues have grabbed some of the market share from ISO order type. More importantly, Figure , Panel B decomposes the growth in off-exchange market share by order types. Midpoint order types grew steadily over time, and are also typically used by institutional traders. Off-exchange trading with larger price improvements "Large PI" closer to midpoint has also grown[18]. Another interesting trend in off-exchange trading is the volatility of "Small PI" trades possibly executed by wholesalers with marginal price improvements.[19]

"Zero PI" is the non-midpoint off-exchange trade with no price improvement (PI), meaning that the executed price is rounded cent. The market share of aggressive trades at bid and offer prices without price improvements is also very volatile but has not grown or declined. Notably, zero

[18]"Large PI" trade is the non-midpoint off-exchange trade with price improvement between 0.4 and 0.6 cent.

[19]The sub-penny price-improvement are identified by Boehmer, Jones, Zhang, and Zhang (2021) as retail trades but their algorithm was challenged by Barber, Huang, Jorion, Odean, and Schwarz (2023) and Battalio, Jennings, Saglam, and Wu (2024) leaving room for further research in this area.

PI trades might allow aggressive traders to find additional market depth in off-exchange venues, without incurring the even higher costs of on-exchange ISOs, according to Chakravarty, Jain, Upson, and Wood (2012).

Table 7 shows the regression results separated by off-exchange trade types that experienced different growth according to Figure . Off-exchange volume for institutional midpoint trades, large PI trades, and aggressive no PI trade type in column 4 increases as the fitted value of on-exchange stock spread increases. The share of Small PI traders in Column 3, identified by small price improvements, do not increase as predicted spread increases. The positive relation between fitted value of spread and market share zero PI trades in Column 4 indicates that informed traders may be using such trades without regard to the high costs because off-exchange venues allow them to trade aggressively in stealth without exposing the orders to market as much as it happens on exchanges. Off-exchange midpoint trades in Column 1 have the strongest long term time-trend of growth but are less sensitive to short term changes in fitted spreads than aggressive traders. The coefficient for fitted value of spreads in Panel B Columns 5 and 6 for on-exchange ISO and non-ISO trades are both negative and significant indicating that traders move from exchange to off-exchange to either avoid higher trading costs or reduce exposure to their informed trading.

The coefficients on the trend variables suggest that the market share of midpoint trades has grown the most with time in our sample. All off-exchange trade types have gained market share and exchange ISO trade types has lost market share in our sample.

4.4.2 Event Days: Regulations, Earnings Announcement and Quad Witching

Off-exchange market share is affected by several factors. In this sections we investigate the impact of regulations, earnings announcements and option expiration on off-exchange market share. Many regulations affect the attractiveness of off-exchange trading venues. On Feb 26, 2007, SEC approved SR-NASD-2005-146, which expanded the scope of IM-2110-2, a rule relative to the minimum price-improvement standard, to apply to OTC equity securities. Based on NASD

Rule 6610(d), OTC equity securities include "exchange-listed securities that do not otherwise qualify for real-time trade reporting". Therefore, off-exchange venues can provide sub-penny price improvement to investors' limit orders after the rule change is approved. This rule increased the attractiveness of off-exchange trading venues and adversely affected exchanges. To deal with the adverse effects of exchanges only receive initial approval for retail price improvement programs in 2012 (Jain, Linna, and McInish (2021)). The exchanges had the goal of recapturing some of the retail trading market share from off-exchange venues through on-exchange sub-penny price improvement. Brogaard and Pan (2022) highlight that the SEC tick size pilot program may affect off-exchange market share. Da, Fang, and Lin (2024); Jain, Mishra, O'Donoghue, and Zhao (2020) show that increased popularity of stock trading among retail investors during Covid remote work arrangement, fractional trading facility offered by brokers, adoption of zero commissions also increased the popularity of off-exchange trading while reducing their informativeness as we see in the last blocks of our Figure 1. Broker Rule changes are beyond the scope of our study but we do control for them in our robustness checks. Time fixed effects also capture the effects of shifts in regulatory regimes.

Our main focus is off-exchange information and price discovery dynamics. We expect off-exchange market share to increase on earnings announcement days, especially for large stocks. We also expect Off-exchange market share to decline on Quad witching days.

Earnings announcement days provide critical information about a firm's financial performance, which can have a significant impact on its stock price. As a result, traders closely monitor these announcements to make their trading decisions on and off-exchange. Balakrishnan, Gkougkousi, Landsman, and Taori (2022) use weekly data to show that dark trading volume increases one week before earnings announcements and reverse after earnings announcement.

In Figure 7, We plot the market share for the period surrounding the firms' earnings announcement period, which includes five days before and after the announcement day. The left panels of the Figure 7 show that off-exchange market share spikes in the period surrounding the

earnings announcement, both in terms of share volume and number of trades. The increase in off-exchange market share begins two days before the earnings announcement and continues to remain high for two days after the announcement. The right panels of the figure separate optionable stocks from the non-optionable ones. The spike in volume and number of trades is much more pronounced for the optionable stocks.

The most evident pattern in Figure 1 was the periodic downward spikes for off-exchange trading volume market share. These significant spikes occur on quad-witching days which are the third Fridays of March, June, September, and December. Quad witching day refers to the simultaneous expiration of four different types of financial instruments: stock options, stock index options, stock index futures, and single stock futures. This suggests that NYSE and NASDAQ grab market share from off-exchange trading on derivative expiration dates. One caveat is that, on quad-witching days, the upward spikes only appear on NYSE-listed securities on NYSE exchange and NASDAQ-listed securities on NASDAQ-Tape C, but not on NASDAQ-Tape A, B. This is because the settlement price for stock-related derivatives is based on listing exchange-traded prices.[20] As a result, arbitrageurs and/or derivative holders prefer to trade on listing lit exchanges rather than other exchanges or off-exchanges to influence the settlement price (Figueiredo, Jain, and Mishra (2023)). This finding along with the earnings spikes document in Figure 7 motivates us to examine additional off-exchange indicator variables for optionable stocks.

Hypothesis 3: Off-exchange market share increases on earnings announcement days and decreases on quad witching derivative expiration days. The relationship in Hypothesis 2 between predicted spread and off-exchange market share should still hold on these event days.

We examine the relationship between off-exchange trading and the predicted quoted spreads with dummy variables separating large firm earnings days, medium and small firm earnings days,

[20]For instance, S&P Global uses real-time exchange-traded prices to calculate intraday S&P DJI index values.More information refer to the document 'S&P DJI's Equity Indices Policies & Practices Methodology' written by S&P Global.

from the rest of the days without any earnings announcements,

$$\begin{aligned}\text{Off-exchange}_{i,t} = &\alpha_i + \beta_1 \widehat{Spread}_{i,t}^{stock} + \beta_2 I_{i,t}^{L-Earnings} + \beta_3 I_{i,t}^{NL-Earnings} \\ &+ \beta_4 I_{i,t}^{L-Witching} + \beta_5 I_{i,t}^{NL-Witching} + \beta_6 Trend_t + Controls_t + \varepsilon_{i,t},\end{aligned} \tag{16}$$

where $I_{i,t}^{L-Earnings}$ and $I_{i,t}^{NL-Earnings}$ are the indicator variables for large and non-large firms' earnings announcements. $I_{i,t}^{L-Witching}$ and $I_{i,t}^{NL-Witching}$ are the quad witching days' indicator variables for large and non-large firms.

Table 8 presents the regression results. The coefficients on predicted stock spreads remain positive and significant consistent with Hypothesis 2. Traders become significantly more active in the off-exchange stock market around earnings announcement for large, medium and small companies (2.1% increase for large firms and 0.7% increase for medium and small firms). Although packed with proportionally more traders off-exchange during earnings announcement days, informed traders in large firms are moving away from exchanges to off-exchange (positive coefficients for information share and density), but for small firms they stay on exchanges for higher execution chances at the firm quotes on-exchange, causing a decline in the off-exchange information density for medium and small firms.

Moving on to Quad Witching Days, our model predicts an increase in off-exchange informed trading as predicted stock spreads rise on exchanges. By including the opposing force of Quad witching in the regression, we test the validity and strength our model predictions despite strong incentives to trade on exchanges. We find that on Quad witching days, off-exchange trading activity spikes down with 10.1% (12.9%) decrease for large firms (medium and small firms) off-exchange market share. The negative coefficients in column 3 imply that some informed traders move to exchanges for faster speed of execution on exchanges, which may be important on quad witching days with a clear expiration deadline. However, information density actually increases by 0.189 for large firms and 0.285 for medium and small firms on quad witching days compared to a regular trading day. indicating the informational role of off-exchange trading as quad witching days pose

a stricter test for the informed traders' choice of off-exchange trading.

Table 9 shows the regression results separated by off-exchange trade types. Consistent with Table 7 results reported previously, off-exchange volume market share for institutional midpoint trades, large PI and aggressive no PI trade types in columns 1, 2, 3 of Table 9 increase and small PI trader in Column 3 do not increase as predicted spread increases. Off-exchange market share increases and exchange market share decreases on earnings announcement days for all trade types within each category for large firms. Off-exchange market share decreases for exchange trade types for medium and small firms also, but results are mixed for the increase in off-exchange market share for medium and small firms. Off-exchange market share on quad witching days declines for all trade types for both large and small firms due to derivative settlement price dynamics, consistent with Hypothesis 3.

4.4.3 Informativeness of Off-exchange Trades

In this section, we test how informed trading intensity (*ITI*) from the machine learning model based on Bogousslavsky, Fos, and Muravyev (2024) affects off-exchange trading volume share and informativeness of off-exchange trades.[21] We estimate three separate regressions based on the following regression:

$$\text{Off-exchange}_{i,t} = \alpha_i + \beta_1 \text{ITI}_{i,t} + \beta_2 \text{Trend}_t + \text{Controls}_t + \varepsilon_{i,t}. \tag{17}$$

Table 10 Panel A reports the results with time trend and control variables to test H2. The positive and statistically significant coefficients on *ITI* of 0.028 in the off-exchange market share regression, 0.018 in off-exchange information share regression, 0.035 in off-exchange information density regression point to the rise of off-exchange informativeness and price discovery. The results can be further interpreted as follows. One standard deviation (0.16; where ITI ranges from 0 to 1) increase

[21] The ITI data is downloaded from Vincent Bogousslavsky's website (https://bogousslavsky.github.io/data/).

in *ITI* leads to 0.45%(0.028×0.16) increase in the off-exchange market share, 0.29% increase in the off-exchange information share, and 0.0056 increase in the off-exchange information density. Thus, with these multi-market dynamics, there is a greater amount of price discovery taking place off-exchange in recent times. Panel A focuses on levels and Panel B focuses on changes in off-exchange market share. We test the role of machine learning based informed trading intensity on the changes in off-exchange volume market share in Panel B. Here again we find that an increase in informed intensity positively and significantly changes the off-exchange market share in model 1. In Model 2, by adding up the coefficients for Stock ITI and Optionable indicator variable, we infer that the impact of information on increasing off-exchange share is stronger for optionable stocks. In Model 3, we observe an even more important result. On Quad witching days, when the settlement rules reduce off-exchange market share (negative coefficient for Witching dummy by itself), the role of Stock ITI (interaction term for Stock ITI× Optionable × Witching) is actually positive.

4.4.4 Sample Expansion: All Common Stocks

Table 11 shows regressions for and expanded sample that includes optionable and non-optionable stocks with indicator and interaction variable for optionable stocks. The purpose of this analysis is to benchmark our model's implication for optionable stocks by including the non-optionable stocks as a control group. We provide the stage 1 coefficients for non-optionable stocks in Table 11 description; they are based on the regression table shown in the internet appendix. Note that option market variables cannot be used to predict stock market spreads for the non-optional group of stocks. For the optionable stocks, we pull the stage 1 coefficients from Table 4 which include option market variables to predict stock market spreads. Stage 2 results in Table 11 confirm the robustness of our findings. Market share and information share increase with the predicted stock market spreads even in this expanded sample, although information density coefficient is insignificant in this combined sample. More important, the positive coefficients for the interaction variable in the 2nd row confirm our model's prediction in Hypothesis 1 that off exchange market share increases

with stock market predicted spreads for optionable stocks. Information share and information density also increases with stock market predicted spreads for optionable stocks. The negative coefficients on the optionable stock indicator variable by itself are consistent with Table 1 summary statistics which showed that the overall off-exchange market share was smaller for optionable stocks than non-optionable stocks because of the on-exchange spikes on quad witching days only for optionable stocks.

5 Off-Exchange Information Predicts Volatility

According to the model explained in Section 2, informed traders initially engage in the options market on day $t-1$ and subsequently in the off-exchange market on day t. Consequently, as this information disseminates over the following days, we expect that the stock's volatility may change in the near future. Thus, our final hypothesis serves as a counterfactual for the null hypothesis that off-exchange trades are uninformed and no impact on price discovery process:

Hypothesis 4: An increase in off-exchange trading volume market share and information share is associated with an increase in realized volatility.

To validate this hypothesis, we employ LASSO on off-exchange and options market data to predict future weekly volatility because LASSO estimation is competitive and beats the HAR lineage, due to higher persistence in the ML models, which helps to approximate the long memory of realized variance. Machine Learning generally increases out-of-sample fit compared to traditional methods (e.g. Chinco, Clark-Joseph, and Ye (2019)) and excels at locating incremental information about future volatility specifically (e.g. Christensen, Siggaard, and Veliyev (2023)).

The objective function of the LASSO model is

$$\min_{\beta} \frac{1}{N} \sum_{(i,t)} (\sigma^2_{i,[t+1,t+5]} - x'_{i,t}\beta)^2 + \lambda \sum_{p=1}^{P} |\beta_p| \tag{18}$$

where $\sigma^2_{[i,t+1,t+5]}$ is the weekly volatility for firm i, $x_{i,t}$ refers to off and on exchange variables for firm i, λ is the penalty coefficient. We use the sum of 5 daily return squared as the proxy for weekly volatility. In order to forecast upcoming weekly volatility, we gather 8 off-exchange and on-exchange variables, 5 options market variables, 6 overall stock market variables, amounting to 19 firm-daily level characteristics.[22] We also add the $t-1$ variables (38 characteristics in total) along with 3 realized variance variables (daily, weekly and monthly realized variance in the HAR-RV model), making a total of 42 predictors.

Our LASSO models are trained using the preceding 750 trading days data (600 training and 150 validation periods), and predict the weekly volatility for the following 22 out of sample trading days (test period). Our training procedure uses a 5-fold cross-validation method which is a common practice in statistics to select the penalty coefficient in LASSO: (1) partition the training set into 5 groups. (2) take out one group as a in-sample validation set. (3) train the model on the remaining data. (4) make predictions on the validation set. This procedure is repeated 5 times (once for each in-sample training set) for a pre-determined penalty coefficient λ. Finally, we select the penalty coefficient with minimum 5-group total mean square error (MSE).

To evaluate our out-of-sample model prediction with our testing periods, we use out-of-sample R^2 measure to evaluate our model performance

$$R^2_{OOS} = 1 - \frac{\sum_{i=1}^{N}(\hat{\sigma}^2_{i,[t+1,t+5]} - \sigma^2_{i,[t+1,t+5]})^2}{\sum_{i=1}^{N}(\bar{\sigma}^2_{i,[t+1,t+5]} - \sigma^2_{i,[t+1,t+5]})^2}. \tag{19}$$

where $\hat{\sigma}_{i,[t+1,t+5]}$ is the model predicted weekly volatility, and $\bar{\sigma}_{i,[t+1,t+5]}$ is the estimated volatility from the benchmark model based on true volatilities $\sigma_{i,[t+1,t+5]}$. We use in-sample cross-sectional mean and predicted volatility from HAR model as benchmark. The positive R^2_{OOS} indicates that the

[22]The 8 off-exchange and on-exchange variables consist of off-exchange market share, information share, information density, on-exchange ISO trade market share, off-exchange small PI trades, midpoint trades, large price improvement trades, and zero price improvement trades market share. The 5 options market variables cover options open interest, options trading volume, options percentage quoted spread, and implied volatilities for at-the-money calls and puts. The 6 overall stock market variables comprise the number of trades, share volume, dollar volume, intraday volatility, percentage quoted spread, and firm size.

model performs better than using either HAR model or cross-sectional mean as predicted volatility.

Table 12 reports the out-of-sample R^2 with our LASSO model relative to different benchmark including sample mean and Heterogeneous Autoregressive model of Realized volatility (HAR) forecasting model (Corsi (2009)) expected, our LASSO model exhibits significantly positive out-of-sample R^2, of 0.31% relative to sample mean with just the off-exchange variables in Model 1. In Model 2, we split off-exchange market share into its 4 different types of order flow. This decomposition further improves the R^2 to 0.59% relative to sample mean and 0.49% relative to HAR. In Model 3, We add firm-specific stock information and options market information. Off-exchange variables identified by us are still useful in forecasting upcoming weekly volatility. Additionally, out-of-sample R^2 rises to 2.39% relative to Sample Mean and 2.29% relative to HAR signifying the improvements with the LASSO model. These LASSO improvements in R^2 with off-exchange predictors are comparable to those in Zhang, Zhang, Cucuringu, and Qian (2024) who use other intra-day indicators to predict volatility.

In Figure 8, we show the heat map of for the LASSO estimation focusing exclusively on 7 off-exchange variables first (Model 2 in Table 12) and then including 39 variables (Model 3 in12). The first 3 years (750 trading days) from 2008 to 2010 is the training set. In model 1, off-exchange variables are selected almost all of the time from 2011 to 2022 as indicated by the dark blue color in the top half. Upon including 39 variables, off-exchange information share, small and large as well as zero and price improvement off-exchange market shares, midpoint off-exchange market share are all still selected in of majority of the validation sample; only the information density variable is unselected in many periods. The bottom two heatmaps present the absolute coefficients in the LASSO models. The color gradients within each column range from highly influential features (dark blue) to the least influential ones (light blue). Small price improvement off-exchange trades provide the most significant incremental information for future volatility prediction, followed by large PI and zero PI off-exchange trades.

6 Conclusion

The last two decades have seen a dramatic increase in off-exchange trading, which now accounts for around 40% of the market. At the trade-by-trade level, a trade in the options market is followed by a dramatic jump in the probability from about 40% to about 60% that the next trade occurs off-exchange. The shift to off-exchange venue by a significant portion of traders persists well beyond the next trade as the off-exchange market share of volume and information share are higher on the next day following the abnormal options activity. Surprisingly, the informativeness of off-exchange trading has also grown, both in terms of traditional measures such as the total amount of information share and information density per unit of trading volume as well as for advanced machine learning methods of detecting informed trading intensity. This growth raises important regulatory and policy questions, as exchanges are subject to both pre-trade and post-trade transparency, while off-exchange trading only provides post-trade information. Our paper provides some insights into the various controversial practices associated with off-exchange trading, as well as the strong industry reaction to proposed regulations.

We develop a two-period model to analyze the behavior of informed traders when selecting among three trading venues: options market, off-exchange, and on-exchange. Our model predicts that in the first period, informed market participants trade aggressively in the options markets, thereby revealing their intentions, causing both options and stock liquidity suppliers to widen the quoted spread in the second period to reduce the risk of adverse selection. This foreseeable defensive strategy forces informed traders to trade off-exchange in the second period.

We conducted an analysis of longitudinal data from 2008 to 2022 to test our stealth trading and cost-based choice mechanism of informed traders. Our cost-based choice mechanism is corroborated by the overall sample, as well as by earning announcement days and major derivative settlement days. We observed both at individual trade level and with daily aggregation that when the predicted quoted spreads are higher, informed traders increasingly opt to trade off-exchange.

Off-exchange trading also rises with the size option trade at the trade level and with open interest at daily aggregation. Around the scheduled earning announcement time, we observe a stronger off-exchange effects for large firms and weaker effects for medium and small firms. During major derivative settlement days, we find a recurring pattern of a decreasing proportion of off-exchange dark trading versus transparent on-exchange trading. Even with this decline, we still find that informed traders choose to trade off-exchange more when the predicted quoted spreads are higher.

The empirical evidence is in line with our model's prediction that when the cost of trading on the exchange rises, the most cost-sensitive uninformed traders move to off-exchange trading. Even those traders who are less cost-sensitive, such as informed traders, switch to off-exchange when the transaction costs reach a certain threshold. Moreover, informed traders can be camouflaged among uninformed traders, particularly when the off-exchange trading volume is high, such as on days when large companies announce their earnings. Consequently, we have seen a significant increase in information shares on earnings announcement days in the post-Covid period (2020-2022). Our findings demonstrate that the off-exchange price discovery is becoming increasingly important as the off-exchange market share grows.

A major implication of our findings is based on our final analysis using machine learning to study how the following weekly volatility changes with off-exchange market shares and information shares. We find that volatility increases as the off-exchange market shares and information shares increases. Thus, the type of informed traders that choose to trade in off-exchange venues increases the volatility of prices in the stock markets.

References

Baillie, R. T., Booth, G. G., Tse, Y., Zabotina, T., 2002. Price discovery and common factor models. Journal of Financial Markets 5, 309–321.

Balakrishnan, K., Gkougkousi, X., Landsman, W. R., Taori, P., 2022. Dark market share around earnings announcements and speed of resolution of investor disagreement. The Accounting Review 97, 1–28.

Bali, T. G., Beckmeyer, H., Moerke, M., Weigert, F., 2023. Option return predictability with machine learning and big data. The Review of Financial Studies 36, 3548–3602.

Barber, B. M., Huang, X., Jorion, P., Odean, T., Schwarz, C., 2023. A (sub) penny for your thoughts: Tracking retail investor activity in taq. Journal of Finance .

Bartlett, R. P., O'Hara, M., 2024. Navigating the murky world of hidden liquidity. Available at SSRN 4988855 .

Battalio, H. R., Jennings, H. R., Saglam, M., Wu, J., 2024. Difficulties in obtaining a representative sample of retail trades from public data sources, *Available at SSRN 4579159*.

Boehmer, E., Jones, C. M., Zhang, X., Zhang, X., 2021. Tracking retail investor activity. Journal of Finance 76, 2249–2305.

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

Brogaard, J., Pan, J., 2022. Dark pool trading and information acquisition. Review of Financial Studies 35, 2625–2666.

Brogaard, J., Ringgenberg, M. C., Roesch, D., 2023. Does floor trading matter? Journal of Finance, forthcoming .

Brugler, J., Comerton-Forde, C., 2022. Differential access to dark markets and execution outcomes. Available at SSRN 3910952 .

Budish, E., Cramton, P., Shim, J., 2015. The high-frequency trading arms race: Frequent batch auctions as a market design response. The Quarterly Journal of Economics 130, 1547–1621.

Buti, S., Rindi, B., Werner, I. M., 2017. Dark pool trading strategies, market quality and welfare. Journal of Financial Economics 124, 244–265.

Buti, S., Rindi, B., Werner, I. M., 2022. Diving into dark pools. Financial Management 51, 961–994.

Chakravarty, S., 2001. Stealth-trading: Which traders' trades move stock prices? Journal of Financial Economics 61, 289–307.

Chakravarty, S., Jain, P., Upson, J., Wood, R., 2012. Clean sweep: Informed trading through intermarket sweep orders. Journal of Financial and Quantitative Analysis 47, 415–435.

Chan, K. F., Marsh, T., 2022. Asset pricing on earnings announcement days. Journal of Financial Economics 144, 1022–1042.

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

Christensen, K., Siggaard, M., Veliyev, B., 2023. A machine learning approach to volatility forecasting. Journal of Financial Econometrics 21, 1680–1727.

Collin-Dufresne, P., Fos, V., 2015. Do prices reveal the presence of informed trading? Journal of Finance 70, 1555–1582.

Comerton-Forde, C., Putniņš, T. J., 2015. Dark trading and price discovery. Journal of Financial Economics 118, 70–92.

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

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Cox, J., 2020. Market fragmentation and post-earnings announcement drift. Journal of Economics and Finance 44, 587–610.

Da, Z., Fang, V. W., Lin, W., 2024. Fractional trading. Review of Financial Studies p. hhae012.

DeHaan, E., Shevlin, T., Thornock, J., 2015. Market (in) attention and the strategic scheduling and timing of earnings announcements. Journal of Accounting and Economics 60, 36–55.

Easley, D., López de Prado, M., O'Hara, M., Zhang, Z., 2021. Microstructure in the machine age. Review of Financial Studies 34, 3316–3363.

Easley, D., O'hara, M., Srinivas, P. S., 1998. Option volume and stock prices: Evidence on where informed traders trade. Journal of Finance 53, 431–465.

Ernst, T., Sokobin, J., Spatt, C., 2021. The value of off-exchange data. In: *Working paper*.

Figueiredo, A., Jain, P., Mishra, S., 2023. The role of fleeting orders on option expiration days. Quantitative Finance 23, 1511–1529.

Hagströmer, B., Menkveld, A. J., 2024. Trades, quotes, and (unbiased) information shares. Available at SSRN 4356262 .

Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. Journal of Finance 50, 1175–1199.

Holden, C. W., Jacobsen, S., 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. Journal of Finance 69, 1747–1785.

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

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

Jain, P. K., Linna, J. A., McInish, T. H., 2021. An examination of the nyse's retail liquidity program. The Quarterly Review of Economics and Finance 80, 367–373.

Jain, P. K., Mishra, S., O'Donoghue, S., Zhao, L., 2020. Trading volume shares and market quality: Pre-and post-zero commissions. Available at SSRN 3741470 .

Menkveld, A. J., Yueshen, B. Z., Zhu, H., 2017. Shades of darkness: A pecking order of trading venues. Journal of Financial Economics 124, 503–534.

Muravyev, D., Pearson, N. D., Broussard, J. P., 2013. Is there price discovery in equity options? Journal of Financial Economics 107, 259–283.

Patton, A. J., Verardo, M., 2012. Does beta move with news? firm-specific information flows and learning about profitability. Review of Financial Studies 25, 2789–2839.

Rahimikia, E., Poon, S.-H., 2020. Machine learning for realised volatility forecasting. Available at SSRN 3707796.

Rosenblatt Securities, 2023. Let there be light. White Paper .

Shkilko, A. V., Van Ness, B. F., Van Ness, R. A., 2008. Locked and crossed markets on nasdaq and the nyse. Journal of Financial Markets 11, 308–337.

Stoll, H. R., 2000. Presidential address: friction. Journal of Finance 55, 1479–1514.

Topbas, Y., Ye, M., 2024. When exchanges are not exchanges: Evidence from two types of dark trading. Available at SSRN .

Xiong, R., Nichols, E. P., Shen, Y., 2015. Deep learning stock volatility with google domestic trends. arXiv preprint arXiv:1512.04916 .

Ye, M., Zhu, W., 2020. Strategic informed trading and dark pools. Available at SSRN 3292516 .

Zhang, C., Zhang, Y., Cucuringu, M., Qian, Z., 2024. Volatility forecasting with machine learning and intraday commonality. Journal of Financial Econometrics 22, 492–530.

Table 1 Average Daily Statistics for Off-Exchange Trading versus Listing Exchanges

This table displays the average daily trading volume in dollars from DTAQ dataset on various exchanges during the sample period from January 1, 2008 to December 31, 2022 for 8,184 exchange-listed firms. We focus on the trades executed during regular trading hours. Off-exchange trades are identified by the exchange code "D", while trades with the exchange code "N" are labeled as NYSE trades. Trades on NASDAQ include codes of "T" and "Q", while all other exchanges are categorized as "Others".

Panel A: Dollar and Share Volume

	Off-exchange			On-exchange Dollar Volume			On-exchange
	Dollar Volume	Share Volume	Market Share	NYSE	NASDAQ	Others	Share Volume
All	\$57.44B	1.68B	37.12%	\$20.87B	\$35.33B	\$52.96B	2.85B
S&P 500	\$40.11B	0.78B	33.19%	\$15.78B	\$25.13B	\$38.28B	1.56B
Non S&P 500	\$17.34B	0.90B	41.34%	\$5.09B	\$10.20B	\$14.69B	1.28B
Optionable	\$56.54B	1.49B	35.84%	\$20.65B	\$19.42B	\$67.70B	2.67B
Non-Optionable	\$0.90B	0.19B	51.98%	\$0.21B	\$0.39B	\$0.79B	0.17B

Panel B: Information Share and Information Density

	Off-exchange		On-exchange	
	Information Share	Information Density	Information Share	Information Density
All	15.47%	0.5859	84.53%	1.3945
Optionable	14.00%	0.4569	86.00%	1.2954
Non-Optionable	24.46%	0.6434	75.54%	1.4032

Panel C: Firm-Level Daily Trading Statistics

	Stock Market			Option Market		
	Mean	Median	Std. Dev	Mean	Median	Std. Dev
Dollar Volume	69.77M	12.39M	299.19M	44.86M	0.37M	690.01M
Percentage Quoted Spread	0.0028	0.0017	0.0037	0.4977	0.4022	0.3275

Table 2 Two-Stage Off-exchange Regressions with Minute-level Aggregation

This table reports the regressions for Amazon with 1, 5, 15, and 30 minute time horizon aggregations. Panel A shows the first stage quoted spread prediction regression:

$$Spread_t^{Stock} = \alpha + \beta_1 \text{Volume}_{t-1}^{Option} + \beta_2 \log(\text{Price})_{t-1} + \text{Fixed Effect}_t + \varepsilon_t.$$

Panel B shows the second stage regression with off-exchange market shares as dependent variable:

$$\text{Off-exchange}_t = \alpha + \beta_1 \widehat{\text{Spread}}_t^{Stock} + \beta_2 \log(\text{Price})_t + \text{Fixed Effect}_t + \varepsilon_{i,t}.$$

The sample period covers from January 1, 2020, to December 31, 2020. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels.

Panel A: Quoted Spread Prediction Regression

	Dollar Quoted Spread				Percentage Quoted Spread			
	1min	5min	15min	30min	1min	5min	15min	30min
$\text{Volume}_{t-1}^{Option} \times 10^{-6}$	26.7***	11.7***	6.29***	3.84***	0.0068***	0.0033***	0.0019***	0.0012***
	(3.24)	(3.97)	(4.78)	(4.63)	(2.43)	(3.32)	(4.25)	(4.13)
$\log(\text{Price})_{t-1}$	0.003	-0.014	0.012	0.026	0.000002	0.000002	0.000014	0.000030
	(0.14)	(-0.36)	(0.20)	(0.30)	(0.25)	(0.17)	(0.62)	(0.09)
Day & Minute Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	97406	19515	6519	3262	97406	19515	6519	3262
Adj R^2	0.7134	0.8127	0.8505	0.8689	0.7159	0.8119	0.8471	0.8676

Panel B: Off-exchange Market Share Regression

	Dollar Quoted Spread				Percentage Quoted Spread			
	1min	5min	15min	30min	1min	5min	15min	30min
$\widehat{\text{Spread}}_t^{Stock}$	0.474***	0.279***	0.148***	0.064**	1861.21***	999.17***	483.8***	207.6**
	(8.76)	(7.80)	(5.60)	(1.98)	(8.75)	(7.85)	(5.54)	(1.99)
$\log(\text{Price})_t$	-0.0004	-0.005	0.002	0.012	-0.002	-0.011*	-0.002	0.013
	(-0.12)	(-0.82)	(0.22)	(0.74)	(-0.78)	(-1.84)	(-0.20)	(0.79)
Day & Minute Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	96700	19402	6488	3248	96700	19402	6488	3248
Adj R^2	0.1516	0.3323	0.4409	0.5037	0.1516	0.3323	0.4409	0.5037

Table 3 Trade-Level Probit Regression

This table reports the probit regressions on Amazon stock trades in Panel A

$$\text{Prob(Off-exchange)} = \alpha + \beta_1 I_{\text{After option}} + \beta_2 I_{\text{Expiration Date}} + \beta_3 \text{Last Quoted Spread} + \text{Controls} + \varepsilon$$

where $I_{\text{After option}}$ is an indicator variable, which takes a value of 1 when the stock trade is preceded by an options trade and 0 otherwise. $I_{\text{Expiration Date}}$ is an indicator variable that equals 1 if the trade date is the expiration date and 0 otherwise. The control variables include the log of the size and price of the previous stock trade, the NBO and NBB depth of the quote prior to the stock trades, the half-hour time interval variable and the day trend variable. The sample period covers from January 1, 2020, to December 31, 2020. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively. Panel B reports the summary statistics of the absolute SHapley Additive exPlanations (SHAP) values for the 19 variables included in the XGBoost model. We compile all stock trades from three trading days before to three trading days after each earnings announcement, yielding a total of 5.36 million observations. Of these, 2.5 million trades are randomly sampled to train the XGBoost model to predict off-exchange execution. The SHAP values are first computed for the remaining 3.36 million trades, and the summary statistics of their absolute values are then presented in Panel B.

Panel A: Probit regression on Amazon stock trades

	(1)	(2)
$I_{\text{After option}}$	0.584***	0.575***
	(1265.00)	(1240.30)
$I_{\text{Expiration Date}}$	-0.040***	-0.039***
	(-84.10)	(-82.81)
Last Quoted Spread	0.147***	0.144***
	(715.53)	(661.50)
Controls	No	Yes
N	45.80M	45.80M

Panel B: Machine Learning Ranked Absolute SHAP Value statistics

	Mean	Max	Median		Mean	Max	Median
$I^{\text{Last Trade}}_{\text{Off-exchange}}$	1.20	2.21	1.24	Last Quote Spread	0.19	2.34	0.14
1s spread change	0.19	1.25	0.18	$I_{\text{After option}}$	0.17	1.22	0.16
Stock trade size	0.10	0.80	0.08	Stock trade price	0.08	0.66	0.07
Option trade price	0.08	0.69	0.07	Minute Indicator	0.05	1.35	0.04
Option Moneyness	0.04	0.70	0.03	5s spread change	0.04	0.53	0.03
Stock NBB	0.04	1.37	0.01	Option trade size	0.04	1.30	0.03
Option quoted spread	0.03	1.27	0.03	Stock NBO	0.03	1.30	0.01
$I_{\text{Expiration Date}}$	0.02	0.37	0.01	Call/Put Indicator	0.02	0.24	0.01
Option NBB	0.02	0.56	0.02	Time to Maturity	0.01	0.63	0.004
Option NBO	0.01	0.48	0.004				

Table 4 The First Stage of Off-exchange Regressions on Optionable Stocks

This table reports the first stage regressions on stock percentage quoted spread for optionable stocks.

$$Spread_{i,t}^{Stock} = \alpha_i + \beta_1 \text{Spread}_{i,t-1}^{Option} + \beta_2 \text{Volatility}_{i,t-1}^{ATM} + \beta_3 \Delta\text{Open Interest}_{i,t-1} + \text{Trend}_t + \text{Controls*}_{t-1} + \varepsilon_{i,t}.$$

The control variables (Controls*$_{t-1}$) include stock percentage quoted spread, log dollar volume, intraday volatility (in percentage), log market capitalization, log price and number of trades. All control variables are lag-1 variables on day $t-1$. The sample period covers from January 1, 2008, to December 31, 2022. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	(1)	(2)	(3)	(4)
$\text{Spread}_{t-1}^{Option}$	0.0008***			0.0003***
	(27.59)			(16.21)
$\text{Volatility}_{i,t-1}^{ATM}$		0.0024***		0.0016***
		(35.76)		(25.80)
$\Delta\text{Open Interest}_{t-1} \times 10^{-6}$			0.0003***	0.0004***
			(3.70)	(3.43)
$\text{Spread}_{t-1}^{Stock}$				0.0922***
				(12.36)
$\log(\text{Volume})_{t-1}^{Stock}$				-0.0001***
				(-4.12)
$\text{Volatility}_{i,t-1}^{Intraday}$				0.0003**
				(2.08)
$\log(\text{MV})_{t-1}^{Stock}$				-0.0005***
				(-8.21)
$\log(\text{Price})_{t-1}^{Stock}$				-0.00001
				(-1.73)
$\log(\text{N})_{t-1}^{Stock}$				-0.0007***
				(-15.81)
Trend	Yes	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
N	8.53M	8.50M	8.41M	7.53M
Adj R^2	0.1952	0.2060	0.2035	0.2349

Table 5 The Second Stage of Off-exchange Regressions on Optionable Stocks

This table reports the regressions for optionable stocks. The first stage regression on stock quoted spread using full sample data

$$Spread_{i,t}^{Stock} = \alpha_i + \beta_1 \text{Spread}_{i,t-1}^{Option} + \beta_2 \text{Volatility}_{i,t-1}^{ATM} + \beta_3 \Delta\text{Open Interest}_{i,t-1} + \text{Trend}_t + \text{Controls}^*_{t-1} + \varepsilon_{i,t}.$$

Estimation results are reported in Table 4. We run the second stage regression on off-exchange measures

$$\text{Off-exchange}_{i,t} = \alpha_i + \beta_1 \widehat{\text{Spread}}_{i,t}^{Stock} + \beta_2 \text{Trend}_t + \text{Controls}_t + \varepsilon_{i,t}.$$

In Panel A, the regression results using the entire sample are presented. In Panel B, we divide our sample into three groups corresponding to the periods 2008-2013, 2014-2019, and 2020-2022, and perform a subsample analysis. The observations with a negative predicted spread are removed. The control variables (Controls$_t$) include SPY off-exchange market share, the market volatility index (VIX), log market capitalization, log price, log number of trades, intraday volatility, log dollar volume, and lag 1 dependent variable. The sample period covers from January 1, 2008, to December 31, 2022. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels.

Panel A: Full sample (2008-2022)

	Market Share		Information Share		Information Density	
$\widehat{Spread}_t^{Stock}$	13.14***	7.13***	21.60***	5.83***	47.14***	4.38***
	(25.53)	(19.53)	(33.69)	(16.45)	(27.54)	(5.43)
Trend$\times 10^{-5}$	-	0.83***	-	1.42***	-	6.97***
	-	(5.08)	-	(8.18)	-	(12.41)
Controls	No	Yes	No	Yes	No	Yes
Firm & Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
N	5.43M	5.43M	4.83M	4.65M	4.83M	4.65M
Adj R^2	0.2945	0.4987	0.1798	0.2107	0.1506	0.1863

Panel B: Subsample Analysis

	Market Share			Information Share			Information Density		
	08-13	14-19	20-22	08-13	14-19	20-22	08-13	14-19	20-22
$\widehat{Spread}_t^{Stock}$	1.54***	7.97***	9.54***	0.49*	5.74***	7.62***	-2.07*	3.29***	6.88***
	(7.63)	(12.87)	(21.64)	(1.85)	(11.95)	(21.01)	(-1.73)	(2.84)	(5.25)
Trend$\times 10^{-5}$	-1.73***	1.21***	3.21***	10.13***	1.94***	0.30	34.60***	9.38***	0.66
	(-5.74)	(4.84)	(12.40)	(25.14)	(8.23)	(1.10)	(26.81)	(11.17)	(0.64)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm & Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1.08M	2.70M	1.65M	0.70M	2.38M	1.57M	0.70M	2.38M	1.57M
Adj R^2	0.4892	0.5230	0.5493	0.1948	0.2176	0.2541	0.0977	0.1968	0.2299

Table 6 Off-exchange LASSO Regressions on Optionable Stocks

This table reports the LASSO regressions for optionable stocks. The objective function of the LASSO model is

$$\min_{\beta} \frac{1}{N} \sum_{(i,t)} (\text{Off-exchange}_{i,t} - x'_{i,t-1}\beta)^2 + \lambda \sum_{p=1}^{P} |\beta_p|$$

where $x_{i,t-1}$ refers to options variables for firm i from the previous day $t-1$, $\text{Spread}^{Option}_{i,t-1}$, $\text{Volatility}^{ATM}_{i,t-1}$, ΔOpen $\text{Interest}_{i,t-1}$, and control variables Controls*_{t-1}, λ is the penalty coefficient. The control variables (Controls*_{t-1}) include stock percentage quoted spread, log dollar volume, intraday volatility (in percentage), log market capitalization, log price and number of trades. To control the autocorrelation of dependent variables, we include lag-1 off-exchange market share, off-exchange information share and off-exchange information density. All control variables are lag-1 variables on day $t-1$. We use out-of-Sample (OOS) R^2 measure to evaluate our model performance

$$R^2_{OOS} = 1 - \frac{\sum_{i=1}^{N} (\widehat{\text{Off-exchange}}_{i,t+1} - \text{Off-exchange}_{i,t+1})^2}{\sum_{i=1}^{N} (Mean(\text{Off-exchange}_{i,t}) - \text{Off-exchange}_{i,t+1})^2}$$

where $\widehat{\text{Off-exchange}}_{i,t+1}$ is the model predicted off-exchange measure, $Mean(\text{Off-exchange}_{i,t})$ is the sample average off-exchange measure from the training set, and $\text{Off-exchange}_{i,t+1}$ is true off-exchange measure. The positive R^2_{OOS} indicates that the LASSO models perform better than the sample average. For each day t, the model is trained on the preceding 750 trading days to validate the model for the following 20 trading days. The data covers January 1, 2008, to December 31, 2022 which yields 150 training windows.

	Market Share (OOS R^2: 0.41%)					Information Share (OOS R^2: 0.33%)				
	Full Sample			2020-2022		Full Sample			2020-2022	
	Mean	Median	Selected %	Mean	Median	Mean	Median	Selected %	Mean	Median
$\text{Spread}^{Option}_{t-1}$	0.0057	0.0055	100.00%	0.0051	0.0051	0.0044	0.0043	100.00%	0.0032	0.0037
$\text{Volatility}^{ATM}_{i,t-1}$	-0.0159	-0.0172	100.00%	0.0038	0.0055	-0.0099	-0.0119	100.00%	0.0112	0.0116
$\Delta\text{OI}_{t-1} \times 10^{-8}$	-0.1450	-0.1500	98.67%	0.0124	0.0136	1.03	0.8670	100.00%	1.94	2.03
$\text{Spread}^{Stock}_{t-1}$	0.1159	0.0533	97.33%	0.4372	0.3275	0.1283	0.1383	99.33%	0.2914	0.1906
$\log(\text{Volume})^{Stock}_{t-1}$	0.0021	0.0015	91.33%	0.0050	0.0055	0.0035	0.0036	100.00%	0.0009	0.0021
$\text{Volatility}^{Intraday}_{i,t-1}$	0.0227	0.0001	95.33%	0.5729	0.5847	0.5443	0.0742	100.00%	0.2971	0.3081
$\log(\text{MV})^{Stock}_{t-1}$	0.0078	0.0079	100.00%	0.0072	0.0070	0.0052	0.0050	100.00%	0.0061	0.0059
$\log(\text{Price})^{Stock}_{t-1}$	-0.0079	-0.0080	100.00%	-0.0078	-0.0084	-0.0077	-0.0082	100.00%	-0.0033	-0.0033
$\log(\text{N})^{Stock}_{t-1}$	-0.0014	-0.0007	100.00%	-0.0062	-0.0070	-0.0027	-0.0029	100.00%	-0.0036	-0.0043

Table 7 Regressions on Optionable Stocks for Various Trade Types

This table reports the regressions on optionable stocks. The first stage regression on stock quoted spread using full sample data

$$Spread_{i,t}^{Stock} = \alpha_i + \beta_1 \text{Spread}_{i,t-1}^{Option} + \beta_2 \text{Volatility}_{i,t-1}^{ATM} + \beta_3 \Delta\text{Open Interest}_{i,t-1} + \text{Trend}_t + \text{Controls}^*_{t-1} + \varepsilon_{i,t}.$$

Estimation results are reported in Table 4. We run the second stage regression on off-exchange measures for various trade types.

$$\text{Trade}_{i,t} = \alpha_i + \beta_1 \widehat{\text{Spread}}_{i,t}^{Stock} + \beta_2 \text{Trend}_t + \text{Controls}_t + \varepsilon_{i,t},$$

where $\text{Trade}_{i,t}$ are various trade types in off-exchange and on-exchange. "Midpoint" trade is the off-exchange trade that its executed price is equal to the midpoint of the NBBO. "Large PI" trade is the non-midpoint off-exchange trade with price improvement greater than 0.4 cent. "Small PI" trade is the off-exchange trade with positive price improvement less than 0.4 cent. "Zero PI" is at the bid or ask with no price improvement (PI). "ISO" is the intermarket sweep order executed on lit-exchange. "Non-ISO" is an on-exchange trade in addition to ISO trades. The control variables (Controls_t) include SPY off-exchange market share, the market volatility index (VIX), log market capitalization, log price, log number of trades, intraday volatility, log dollar volume, and lag 1 dependent variable. The sample period covers from January 1, 2008, to December 31, 2022. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	Off-exchange				On-exchange	
	Midpoint	Large PI	Zero PI	Small PI	ISO	Non-ISO
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{Spread}_t^{Stock}$	0.99***	5.56***	1.17***	-0.78***	-0.79***	-6.26***
	(14.41)	(21.00)	(11.50)	(-10.34)	(-6.41)	(-20.54)
Trend$\times 10^{-5}$	1.21***	0.88***	-1.24***	0.18***	-4.60***	4.07***
	(23.73)	(8.13)	(-17.84)	(4.29)	(-49.38)	(30.85)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
N	5.43M	5.43M	5.43M	5.43M	5.43M	5.43M
Adj R^2	0.1926	0.4501	0.4573	0.6935	0.4368	0.4499

Table 8 Off-exchange Regressions on Optionable Stocks for Earnings Announcement Days and Quad Witching Days

This table reports the regressions on optionable stocks. The first stage regression on stock quoted spread using full sample data

$$Spread_{i,t}^{Stock} = \alpha_i + \beta_1 \text{Spread}_{i,t-1}^{Option} + \beta_2 \text{Volatility}_{i,t-1}^{ATM} + \beta_3 \Delta\text{Open Interest}_{i,t-1} + \text{Trend}_t + \text{Controls}^*_{t-1} + \varepsilon_{i,t}.$$

Estimation results are reported in Table 4. We run the second stage regression on off-exchange measures for all trading days.

$$\begin{aligned}\text{Off-exchange}_{i,t} = {} & \alpha_i + \beta_1 \widehat{Spread}_{i,t}^{stock} + \beta_2 I_{i,t}^{L-Earnings} + \beta_3 I_{i,t}^{NL-Earnings} \\ & + \beta_4 I_{i,t}^{L-Witching} + \beta_5 I_{i,t}^{NL-Witching} + \beta_6 Trend_t + Controls_t + \varepsilon_{i,t},\end{aligned}$$

where $I_{i,t}^{L-Earnings}$ and $I_{i,t}^{NL-Earnings}$ are the indicator variables for large and non-large firms' earnings announcements. $I_{i,t}^{L-Witching}$ and $I_{i,t}^{NL-Witching}$ are the quad witching days' indicator variables for large and non-large firms. The control variables (Controls_t) include SPY off-exchange market share, the market volatility index (VIX), log market capitalization, log price, log number of trades, intraday volatility, log dollar volume, and lag 1 dependent variable. The sample period covers from January 1, 2008, to December 31, 2022. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	Market Share		Information Share		Information Density	
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{Spread}_t^{Stock}$	7.03***	7.01***	5.64***	7.68***	3.76***	10.15***
	(19.34)	(19.30)	(16.21)	(21.04)	(4.67)	(12.19)
$I_{i,t}^{L-Earnings}$	0.021***	0.022***	0.030***	0.031***	0.073***	0.069***
	(36.14)	(36.10)	(32.93)	(31.40)	(23.27)	(20.20)
$I_{i,t}^{NL-Earnings}$	0.007***	0.009***	0.004***	0.004***	0.006**	-0.002
	(15.11)	(17.41)	(4.95)	(4.63)	(2.13)	(-0.77)
$SUR \times I_{i,t}^{L-Earnings}$		0.004***		0.016		0.163**
		(3.50)		(0.92)		(2.56)
$SUR \times I_{i,t}^{NL-Earnings}$		-0.0007		0.007*		0.043***
		(-0.26)		(1.81)		(3.03)
$I_{i,t}^{L-Witching}$	-0.101***	-0.101***	-0.008***	-0.006***	0.189***	0.188***
	(-77.24)	(-77.23)	(-11.93)	(-10.00)	(31.81)	(31.47)
$I_{i,t}^{NL-Witching}$	-0.129***	-0.129***	-0.006***	-0.005***	0.285***	0.295***
	(-92.65)	(-92.64)	(-8.31)	(-6.01)	(46.05)	(46.73)
Trend$\times 10^{-5}$	1.95***	1.96***	1.57***	1.15***	4.74***	4.44***
	(11.85)	(11.93)	(9.06)	(6.08)	(8.30)	(7.10)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
N	5.43M	5.42M	4.65M	4.65M	4.65M	4.65M
Adj R^2	0.5136	0.5133	0.2111	0.2031	0.1895	0.1788

Table 9 Regressions on Optionable Stocks around Earnings Announcement Days and Quad Witching Days for Various Trade Types

This table reports the regressions on optionable stocks for EADs. The first stage regression on stock quoted spread using all optionable stocks data

$$\begin{aligned} Spread_{i,t}^{Stock} =& \alpha_i + \beta_1 \text{Spread}_{i,t-1}^{Option} + \beta_2 \text{Volatility}_{i,t-1}^{ATM} + \beta_3 \Delta \text{Open Interest}_{t-1} \\ &+ \beta_4 \text{Spread}_{i,t-1}^{Stock} + \beta_5 \text{Volatility}_{i,t-1}^{Intraday} + \beta_6 \text{Volume}_{t-1}^{Stock} + \text{Trend}_t + \varepsilon_{i,t}. \end{aligned}$$

Estimation results are reported in Table 4 Panel A. We run the second stage regression on off-exchange measures for various trade types.

$$\begin{aligned} \text{Trade}_{i,t} =& \alpha_i + \beta_1 \widehat{Spread}_{i,t}^{stock} + \beta_2 I_{i,t}^{L-Earnings} + \beta_3 I_{i,t}^{NL-Earnings} \\ &+ \beta_4 I_{i,t}^{L-Witching} + \beta_5 I_{i,t}^{NL-Witching} + \beta_6 Trend_t + Controls_t + \varepsilon_{i,t}, \end{aligned}$$

where $I_{i,t}^{L-Earnings}$ and $I_{i,t}^{NL-Earnings}$ are the indicator variables for large and non-large firms' earnings announcements. $I_{i,t}^{L-Witching}$ and $I_{i,t}^{NL-Witching}$ are the quad witching days' indicator variables for large and non-large firms. The control variables (Controls_t) include SPY off-exchange market share, the market volatility index (VIX), log market capitalization, log price, log number of trades, intraday volatility, log dollar volume, and lag 1 dependent variable. The sample period covers from January 1, 2008, to December 31, 2022. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	Off-exchange				On-exchange	
	Midpoint	Large PI	Zero PI	Small PI	ISO	Non-ISO
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{Spread}_t^{Stock}$	1.00***	5.36***	1.27***	-0.80***	-0.71***	-6.22***
	(14.36)	(20.78)	(12.09)	(-10.57)	(-5.82)	(-20.37)
$I_{i,t}^{L-Earnings}$	0.0026***	0.0280***	-0.0082***	0.0012***	-0.0095***	-0.0123***
	(9.23)	(64.59)	(-24.35)	(4.82)	(-22.95)	(-26.84)
$I_{i,t}^{NL-Earnings}$	-0.0013***	0.0178***	-0.0086***	0.0019***	-0.0066***	-0.0023***
	(-4.90)	(45.46)	(-25.53)	(8.13)	(-15.86)	(-5.30)
$I_{i,t}^{L-Witching}$	-0.0221***	-0.0408***	-0.0309***	-0.0077***	-0.0289***	0.1298***
	(-61.14)	(-64.48)	(-64.08)	(-46.30)	(-45.17)	(72.73)
$I_{i,t}^{NL-Witching}$	-0.0217***	-0.0448***	-0.0441***	-0.0176***	-0.0334***	0.1612***
	(-52.26)	(-66.08)	(-76.93)	(-59.43)	(-56.84)	(91.88)
$\text{Trend} \times 10^{-5}$	1.41***	1.36***	-0.92***	0.39***	-4.33***	2.70***
	(26.91)	(12.66)	(-13.04)	(4.31)	(-46.74)	(20.34)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
N	5.42M	5.42M	5.42M	5.42M	5.42M	5.42M
Adj R^2	0.1954	0.4558	0.4615	0.6944	0.4387	0.4830

Table 10 Off-exchange Regressions on Informed Trading Intensity (ITI)

This table reports the regressions on off-exchange measures

$$\text{Off-exchange}_{i,t} = \alpha_i + \beta_1 \text{ITI}_{i,t} + \beta_2 \text{Trend}_t + \text{Controls}_t + \varepsilon_{i,t}.$$

where $\text{ITI}_{i,t}$ is the informed trading intensity based on the machine learning method described in Bogousslavsky, Fos, and Muravyev (2024) for firm i on day t. The control variables (Controls_t) include SPY off-exchange market share, the market volatility index (VIX), log market capitalization, log price, log number of trades, intraday volatility, log dollar volume, and lag 1 dependent variable. The sample period covers from January 1, 2008, to July 31, 2021. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	Off-Exchange Market Share	Off Exchange Information Share	Off-Exchange Information Density
	(1)	(2)	(3)
Stock $\text{ITI}_{i,t}$	0.028***	0.018***	0.035***
	(38.89)	(20.68)	(10.57)
Trend$\times 10^{-4}$	0.23***	0.42***	1.18***
	(21.30)	(36.67)	(25.85)
Controls	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
N	6.55M	4.92M	4.92M
Adj R^2	0.3861	0.2561	0.2037
	Change of Off-Exchange Market Share		
	(1)	(2)	(3)
Stock $\text{ITI}_{i,t}$	0.022***	0.016***	0.012***
	(43.22)	(10.49)	(8.16)
Stock $\text{ITI}_{i,t} \times I_{i,t}^{\text{Optionable}}$		0.009***	0.012***
		(5.04)	(7.39)
Stock $\text{ITI}_{i,t} \times I_{i,t}^{\text{Optionable}} \times I_t^{\text{Witching}}$			0.008***
			(2.96)
$I_{i,t}^{\text{Optionable}}$		0.002***	0.002***
		(3.03)	(3.48)
I_t^{Witching}			-0.105***
			(-70.44)
Trend$\times 10^{-4}$	0.18***	0.18***	0.25***
	(36.36)	(36.73)	(45.95)
Controls	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
N	6.55M	6.55M	6.55M
Adj R^2	0.0520	0.0521	0.0596

Table 11 Off-Exchange Activity Regressions for All Stocks

This table reports the regressions with the expanded sample of all stocks. For optionable stocks, the first stage regression on stock quoted spread as

$$Spread_{i,t}^{Stock} = \alpha_i + \beta_1 \text{Spread}_{i,t-1}^{Option} + \beta_2 \text{Volatility}_{i,t-1}^{ATM} + \beta_3 \Delta\text{Open Interest}_{i,t-1} + \text{Trend}_t + \text{Controls}^*_{t-1} + \varepsilon_{i,t},$$

where the control variables (Controls$^*_{t-1}$) include stock percentage quoted spread, log dollar volume, intraday volatility, log market capitalization, log price and number of trades. All control variables are lag-1 variables on day $t-1$. Estimation results are reported in Table 4. The first stage regression for non-optionable stocks is based on Stoll (2000), where the spread is regressed on control variables along with firm-specific effects.

$$\begin{aligned} Spread_{i,t}^{Stock} = &\alpha_i + 0.22\text{Spread}_{i,t-1}^{Stock} - 0.002\log(\text{Volume})_{i,t-1}^{Stock} + 0.2319\text{Volatility}_{i,t-1}^{Intraday} - 0.0061\log(\text{MV})_{i,t-1}^{Stock} \\ &+ 0.0002\log(\text{Price})_{i,t-1}^{Stock} - 0.0011\log(\text{N})_{i,t-1}^{Stock}. \end{aligned}$$

We report the second stage regression on off-exchange measures

$$\text{Off-exchange}_{i,t} = \alpha_i + \beta_1 \widehat{Spread}_{i,t}^{stock} + \beta_2 I_{i,t}^{Optionable} \widehat{Spread}_{i,t}^{stock} + \beta_3 I_{i,t}^{Optionable} + \beta_4 \text{Trend}_t + \text{Controls}_t + \varepsilon_{i,t}.$$

where $I_{i,t}^{Optionable}$ is the indicator variables for optionable stocks. We only include firm-day observations with price greater than 1 and number of trades greater than 100. The control variables (Controls$_t$) include SPY off-exchange market share, the market volatility index (VIX), log market capitalization, log price, log number of trades, intraday volatility, log dollar volume, and lag 1 dependent variable. The sample period covers from January 1, 2008, to December 31, 2022. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	Market Share (1)	Information Share (2)	Information Density (3)
$\widehat{Spread}_t^{Stock}$	1.04***	0.75***	-0.30
	(20.94)	(9.44)	(-1.39)
$\widehat{Spread}_{i,t}^{stock} \times I_{i,t}^{Optionable}$	3.09***	4.97***	5.02***
	(16.40)	(18.69)	(7.83)
$I_{i,t}^{Optionable}$	-0.02***	-0.05***	-0.08***
	(-12.14)	(-27.04)	(-13.84)
Trend$\times 10^{-5}$	0.41***	1.16***	6.27***
	(3.44)	(7.10)	(11.56)
Control	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
N	7.15M	5.14M	5.14M
Adj R^2	0.5114	0.2264	0.1862

Table 12 Out-of-sample performance of weekly volatility predication with off-exchange variables using machine learning

This table presents the out-of-sample performance of LASSO models that leverages both off-exchange and options market data to forecast upcoming weekly volatility. The objective function of the LASSO model is

$$\min_{\beta} \frac{1}{N} \sum_{(i,t)} (\sigma^2_{i,[t+1,t+5]} - x'_{i,t}\beta)^2 + \lambda \sum_{p=1}^{P} |\beta_p|$$

where $\sigma^2_{[i,t+1,t+5]}$ is the weekly volatility for firm i, $x_{i,t}$ refers to off and on exchange variables for firm i, λ is the penalty coefficient. We use the sum of 5 daily return squared as the proxy for weekly volatility. The 3 off-exchange variables refer to off-exchange market share, off-exchange information share, off-exchange information density. The 7 off-exchange and on-exchange variables consist of off-exchange information share, information density, on-exchange ISO trade market share, off-exchange zero, small and large price improvement trades market shares, midpoint trades. The 5 options market variables cover options open interest, options trading volume, options percentage quoted spread, and implied volatilities for at-the-money calls and puts. The 6 overall stock market variables comprise the number of trades, share volume, dollar volume, intraday volatility, percentage quoted spread, and firm size, amounting to 18 total daily level characteristics. According to our model, we use LASSO to predict future weekly volatility employing data from the last two days (36 characteristics) along with 3 past variance, making a total of 39 predictors. We use out-of-sample R^2 measure to evaluate our model performance

$$R^2_{OOS} = 1 - \frac{\sum_{i=1}^{N} (\hat{\sigma}^2_{i,[t+1,t+5]} - \sigma^2_{i,[t+1,t+5]})^2}{\sum_{i=1}^{N} (\bar{\sigma}^2_{i,[t+1,t+5]} - \sigma^2_{i,[t+1,t+5]})^2}$$

where $\hat{\sigma}^2_{i,[t+1,t+5]}$ is the model predicted weekly volatility, and $\bar{\sigma}^2_{i,[t+1,t+5]}$ is the estimated volatility from the benchmark model based on true volatilities $\sigma^2_{i,[t+1,t+5]}$. The positive R^2_{OOS} indicates that the LASSO models perform better than the benchmark models (sample average and HAR model). For each day t, the model is trained on the preceding 750 trading days to predict the weekly volatility for the following 20 trading days. The data covers January 1, 2008, to December 31, 2022.

		1-day R^2_{OOS} relative to		5-day R^2_{OOS} relative to	
Model	Variable Sets	Sample Mean	HAR	Sample Mean	HAR
1 (3 Variables)	Off-exchange Market Share, Information Share, Information Density	0.09%	0.06%	0.31%	0.21%
2 (7 Variables)	Off-exchange Information Share, Information Density, Four Off-exchange Order Flow Market Shares, On-exchange ISO Market Shares	0.16%	0.13%	0.59%	0.49%
3 (39 Variables)	Model 2+ Stock Information, Options Information and Realized Variance	0.90%	0.78%	2.39%	2.29%

Figure 1. Time Series of Off-exchange Market Share and Information Share

The figures show the time series of off-exchange market share and information share. Off-exchange trades are identified by the exchange code "D". The sample period covers January 1, 2008, to December 31, 2022.

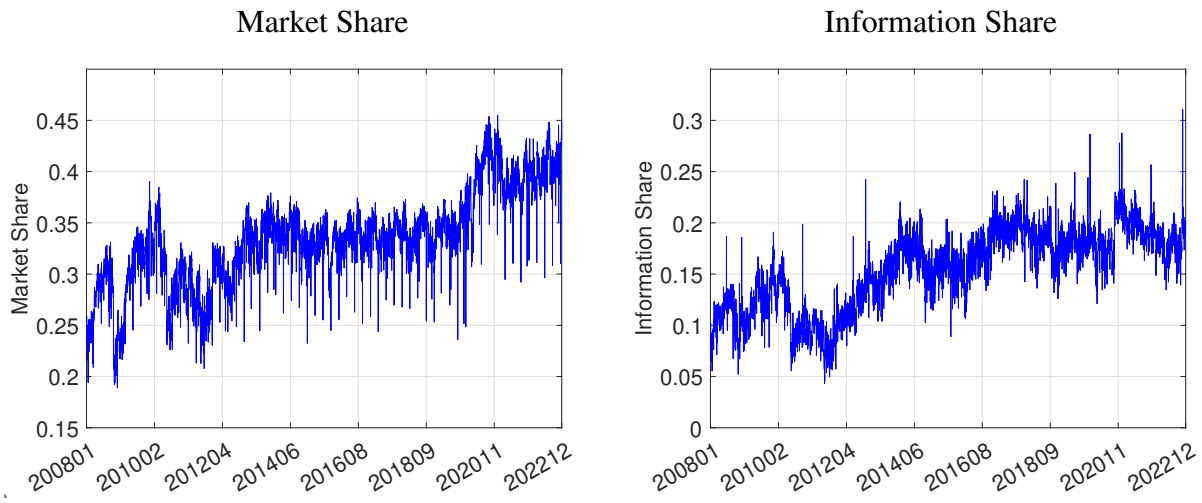


Figure 2. Road Map of the Two-Period Model

Informed traders reveal their intentions in the options market in period $t-1$, and the stock liquidity suppliers' defensive strategy is predictable and it forces informed traders to trade in off-exchange in period t. Therefore, both Off-exchange market shares (MS) and information share (IS) increase.

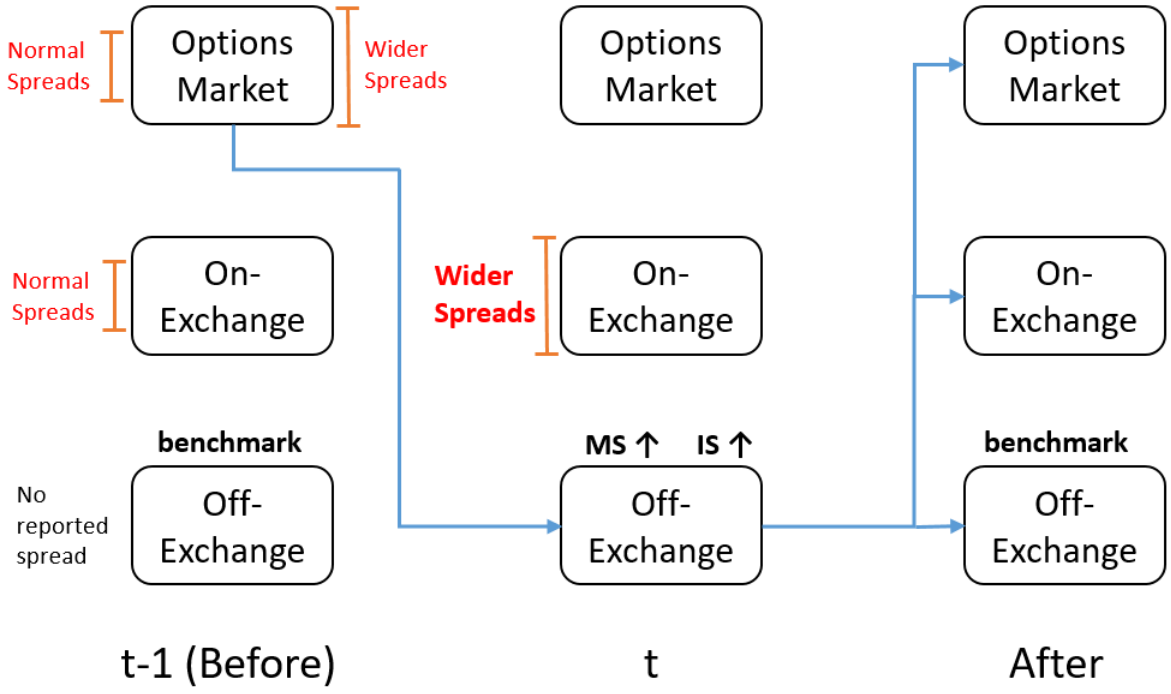


Figure 3. Dollar Quoted Spread Change Before and After the Amazon Option Trade
The figures show the change in the average dollar quoted spread surrounding Amazon option trades. We collect the quoted spread at 100 millisecond intervals from five seconds before to five seconds after each trade, and compute the average based on the 25.02 million option trades executed in 2020. The sample period covers from January 1, 2020, to December 31, 2020.

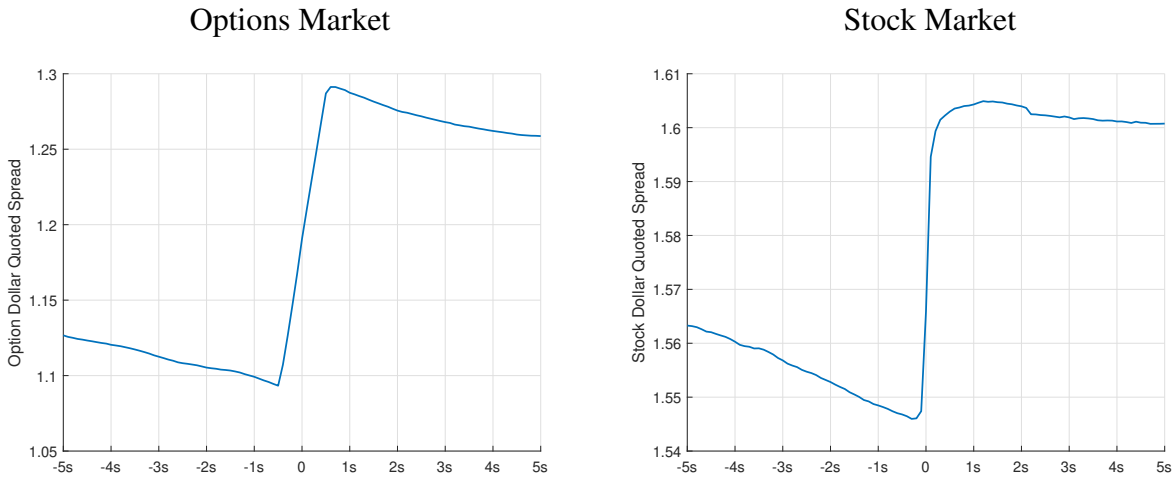


Figure 4. Time Series of Intraday Probability of Off-exchange Trade

The figure presents the unconditional intraday probability of off-exchange trading and the conditional intraday probability of off-exchange trading following the execution of an options trade, focusing exclusively on Amazon stock and trades. The unconditional probability is calculated as the ratio of off-exchange trades to total trades, while the conditional probability is determined based on the first stock trade occurring after an options trade. The sample period covers January 1, 2020, to December 31, 2020.

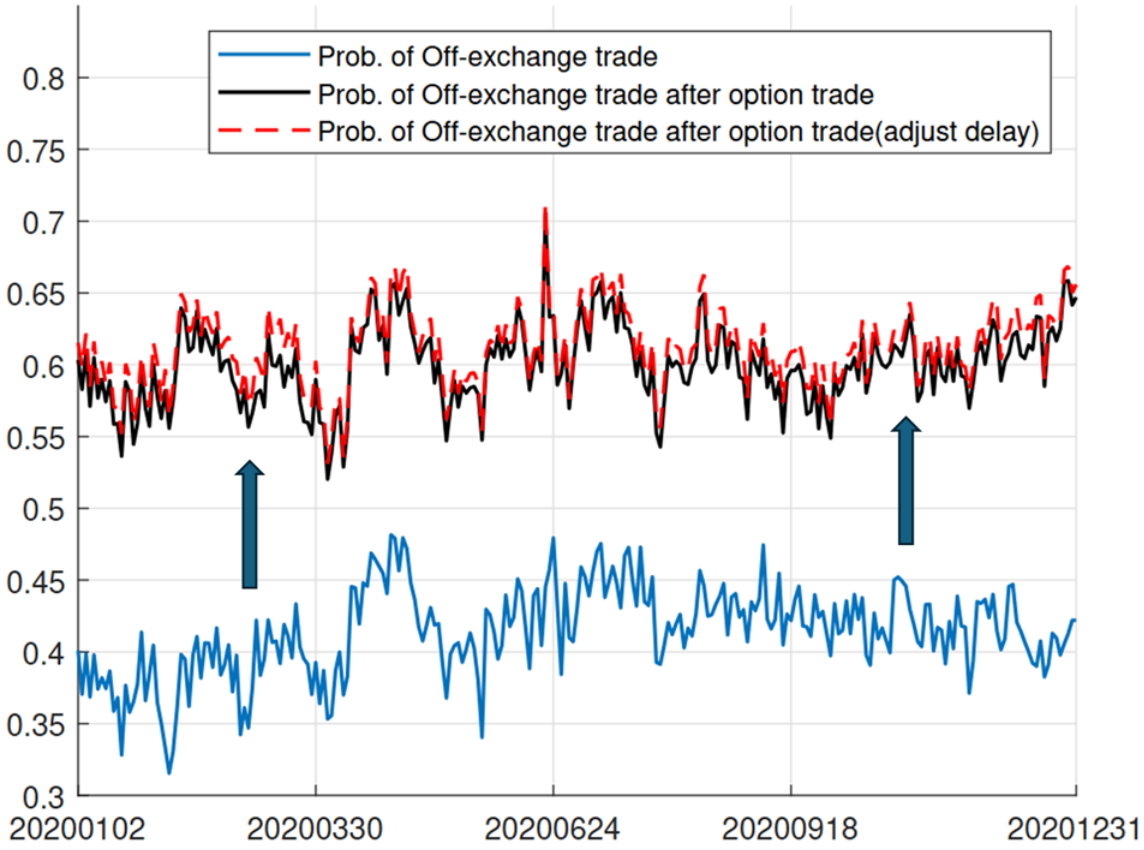


Figure 5. SHAP Value Scatter Plot for Determinants of Off-Exchange Trade Execution
The figure plots the SHAP values against the underlying raw data for two variables: the after option trade dummy and the stock quoted spread. We compile all stock trades from three trading days before to three trading days after each earnings announcement, yielding a total of 5.36 million observations. Of these, 2.5 million trades are randomly sampled to train the XGBoost model to predict off-exchange execution, and SHAP values are computed for the remaining 3.36 million trades.

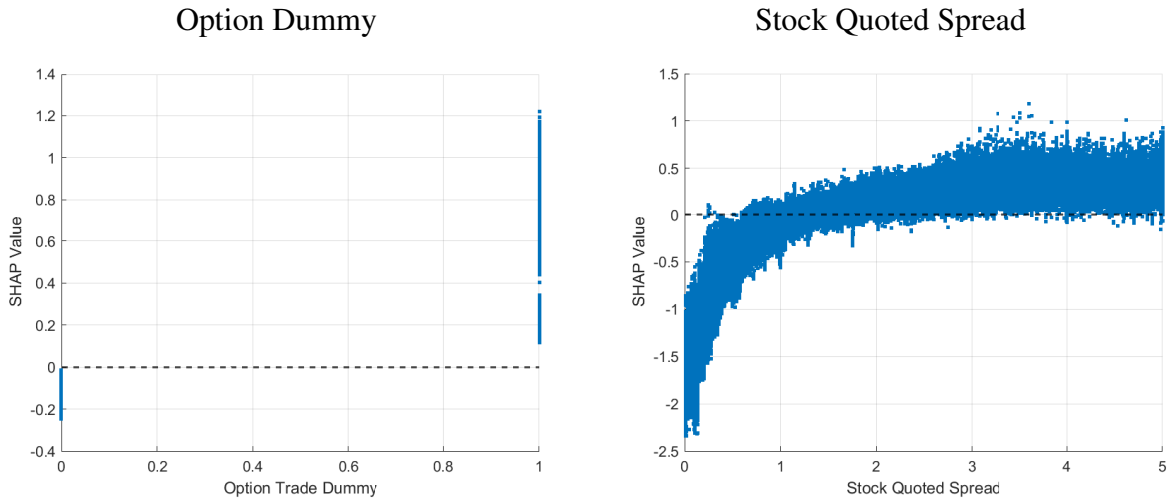


Figure 6. Decomposition of On-exchange and Off-exchange Market Share

The figures depict time series of on-exchange and off-exchange market share for various types of trades. "ISO" trade is the intermarket sweep order executed on-exchange. "Midpoint" trade is the off-exchange trade that its executed price is equal to the midpoint of the NBBO. "Large PI" trade is the non-midpoint off-exchange trade with price improvement greater than 0.4 cent. "Small PI" trade is the off-exchange trade with positive price improvement less than 0.4 cent. "Zero PI" is at the bid or ask with no price improvement (PI). The sample period covers from January 1, 2008, to December 31, 2022.

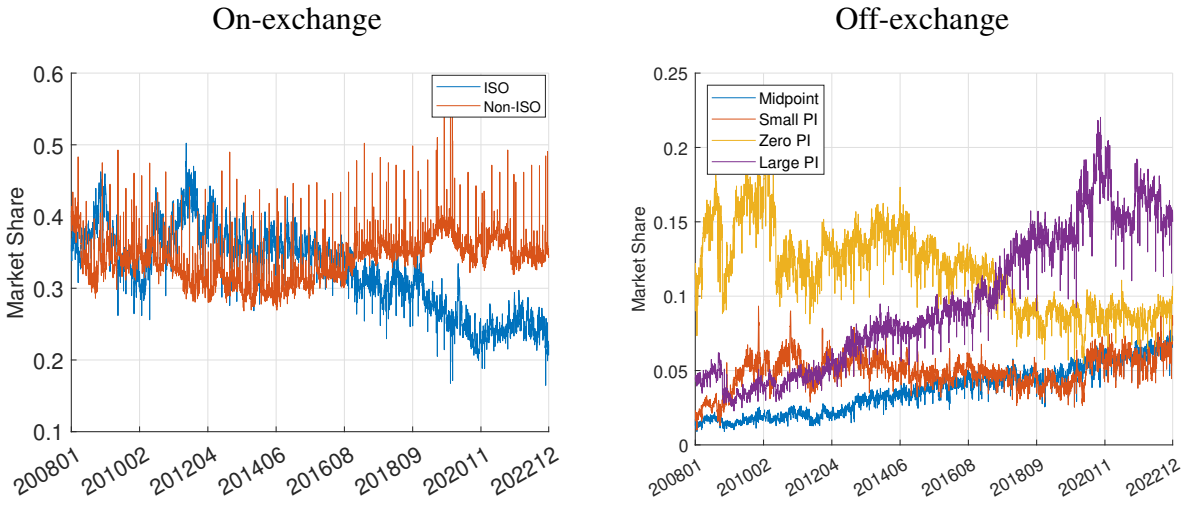


Figure 7. Market Shares of Off-exchange Trading around Earnings Announcement Days
The figures present the off-exchange market shares around earnings announcement days (EADs). The first and second rows of figures present the market shares based on share volume measures and number of trade measures, respectively. The sample period covers January 1, 2008, to December 31, 2022.

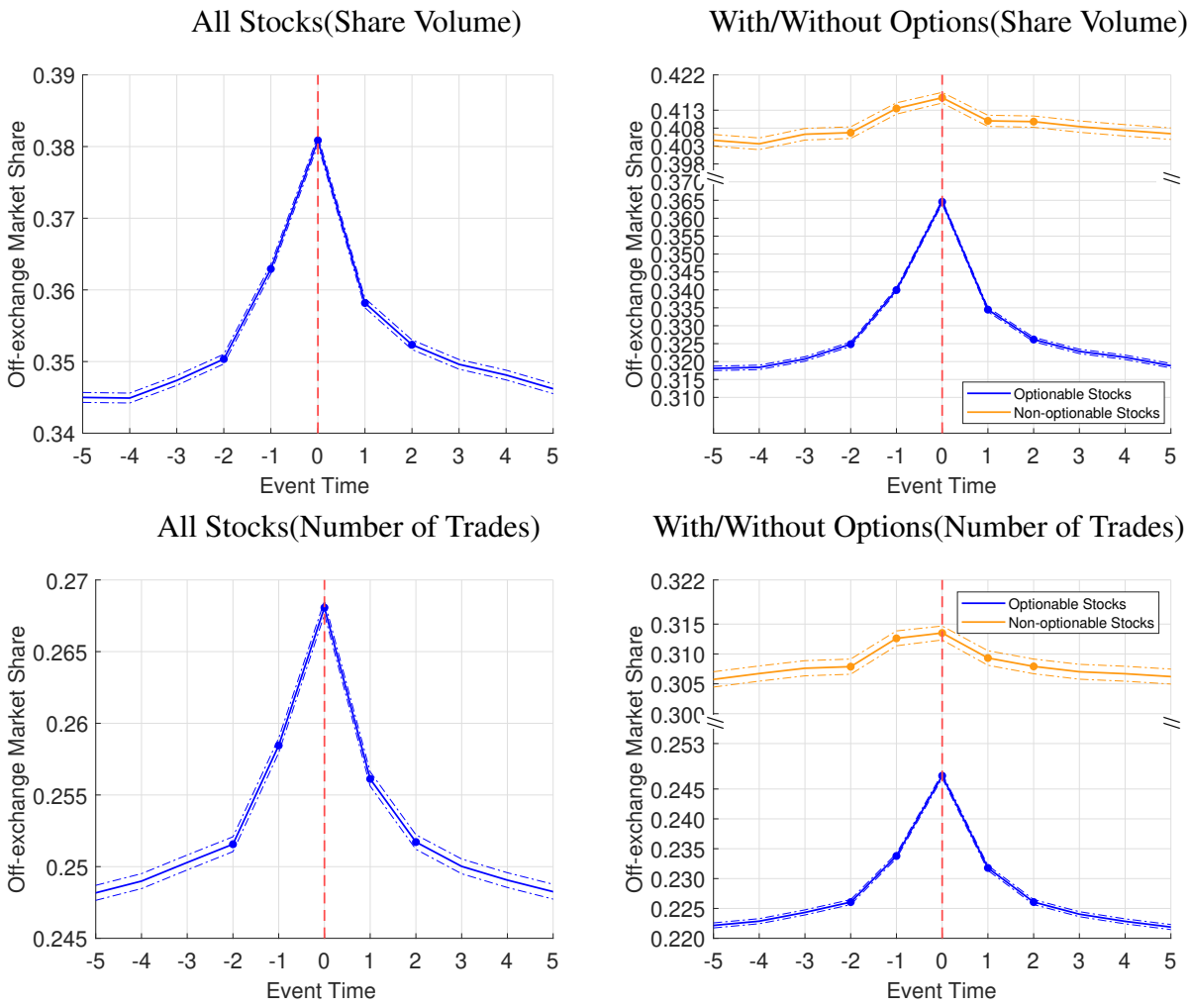
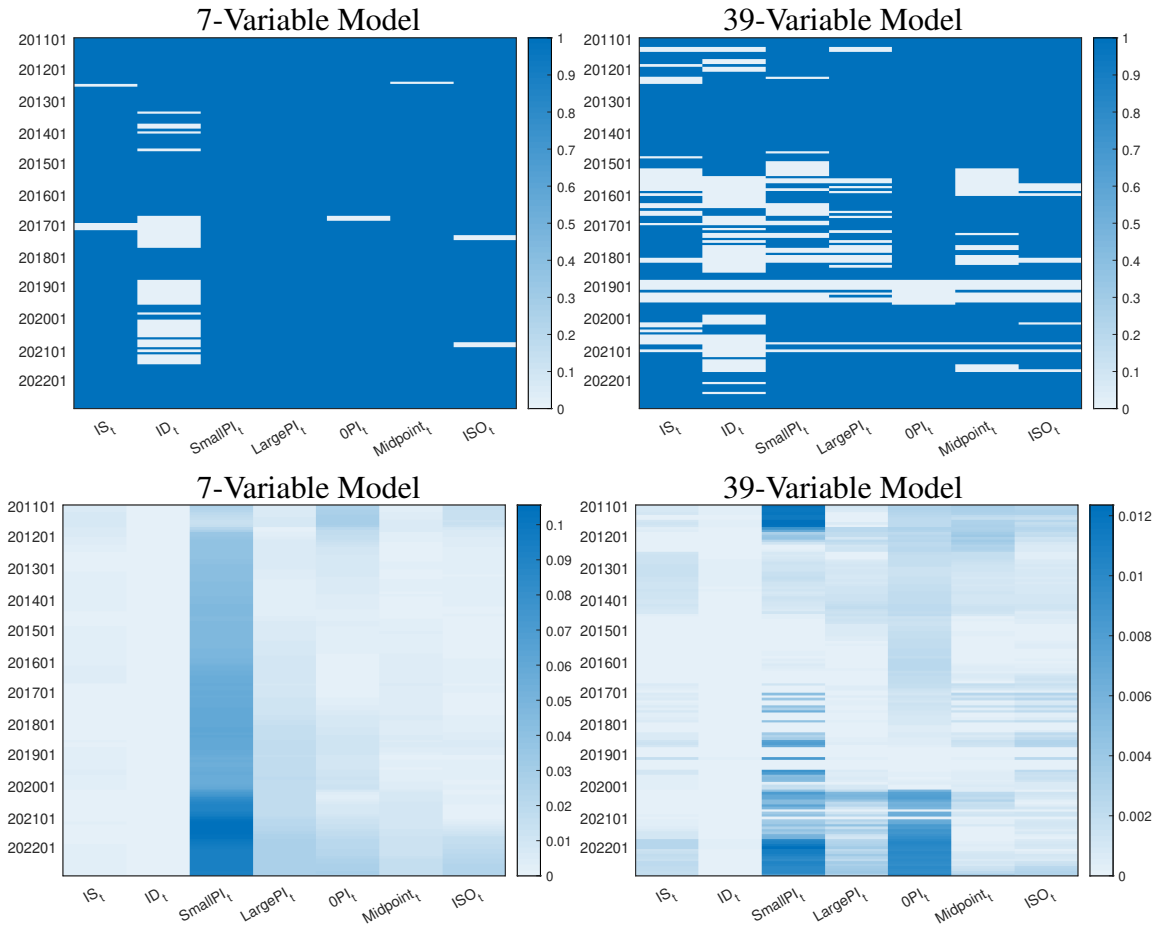


Figure 8. Heatmaps of LASSO Models

The figure presents the heatmap of two LASSO models. We mainly focus on 7 variables which include off-exchange information share (IS_t), information density (ID_t), four off-exchange order flows market shares ($SmallPI_t$, $LargePI_t$, $0PI_t$ and $Midpoint_t$) and on-exchange ISO market share (ISO_t). The top two heatmaps illustrate the variable selection in the LASSO models, where dark blue indicates a selected variable and white signifies an unselected one. The bottom two heatmaps present the absolute coefficients in the LASSO models. The color gradients within each column range from highly influential features (dark blue) to the least influential ones (light blue). The sample period covers January 1, 2008, to December 31, 2022. For each day t, the model is trained on the preceding 750 trading days to predict the weekly volatility for the following days, therefore the heat map starts from January 1, 2011 to December 31, 2022.



Appendix

A1 Exchange ID in Daily TAQ database

Each trade and quote message in the daily TAQ database contains a participant ID, which serves to identify the exchange or processor that initiated the message.

Table A1 List of Exchange ID

ID	Description
Panel A: NYSE Group of Exchanges	
N	New York Stock Exchange, LLC (NYSE)
P	NYSE Arca, Inc. (NYSE Arca)
A	NYSE American, LLC (NYSE American)
C	NYSE National, Inc (NYSE National)
M	Chicago Stock Exchange, Inc. (NYSE Chicago)
Panel B: Nasdaq Group of Exchanges	
T	NASDAQ Stock Exchange, LLC (in Tape A, B securities)
Q	NASDAQ Stock Exchange, LLC (in Tape C securities)
B	NASDAQ OMX BX, Inc. (NASDAQ OMX BX)
X	NASDAQ OMX PSX, Inc. (NASDAQ OMX PSX)
I	International Securities Exchange, LLC (ISE)
Panel C: CBOE Group of Exchanges	
W	Chicago Broad Options Exchange, Inc.(CBSX)
Y	Cboe BYX Exchange, Inc (Cboe BYX)
Z	Cboe BZX Exchange, Inc (Cboe BZX)
J	Cboe EDGA Exchange (Cboe EDGA)
K	Cboe EDGX Exchange (Cboe EDGX)
Panel D: Independent Venues	
D	Financial Industry Regulatory Authority, Inc. (FINRA ADF)
V	The Investors' Exchange, LLC (IEX)
L	Long-Term Stock Exchange (LTSE)
H	MIAX Pearl, LLC (MIAX)
U	Members Exchange (MEMX)

A2 Descriptive Statistics of Optionable Stocks

Table A2 Descriptive Statistics of Optionable Stocks

	All (N=2312)			Large Firms (N=1214)			Medium Firms (N=854)			Small Firms (N=245)		
	Mean	Median	Std Dev.	Mean	Median	Std Dev.	Mean	Median	Std Dev.	Mean	Median	Std Dev.
Panel A: All trading days												
Size (Billions)	9.39	1.59	37.72	17.84	4.93	51.60	0.74	0.67	0.37	0.12	0.12	0.05
Price ($)	42.90	26.47	67.74	67.99	48.31	87.19	20.62	16.16	16.33	5.05	4.00	3.75
Stock Volume												
Number of Trades	10,474	4,319	21,408	16,839	9,025	27,225	3,915	2,299	6,417	1,614	786	3,660
Share Volume ($\times 10^6$)	1.83	0.50	6.76	2.80	0.97	8.49	0.69	0.25	2.20	0.48	0.14	1.76
Dollar Volume ($\times 10^6$)	69.77	12.39	299.19	128.26	43.54	408.22	8.07	4.19	17.72	1.50	0.59	4.68
Off-exchange Volume												
Number of Trades	2,472	887	6,553	3,922	1,772	8,474	979	478	2,245	530	217	1,490
Share Volume ($\times 10^6$)	0.65	0.15	2.95	0.96	0.27	3.63	0.28	0.07	1.15	0.25	0.06	1.01
Dollar Volume ($\times 10^6$)	23.72	3.58	126.68	43.79	12.59	175.07	3.00	1.26	8.67	0.72	0.24	2.60
On-exchange Volume												
Number of Trades	8,002	3,394	15,172	12,917	7,203	19,145	2,936	1,797	4,288	1,083	553	2,215
Share Volume ($\times 10^6$)	1.18	0.34	3.90	1.84	0.69	4.96	0.41	0.17	1.09	0.23	0.08	0.76
Dollar Volume ($\times 10^6$)	46.06	8.54	175.64	84.47	30.38	237.43	5.08	2.83	9.41	0.77	0.33	2.14
Percentage Quoted Spread												
Stock Market	0.0028	0.0017	0.0037	0.0012	0.0009	0.0010	0.0034	0.0028	0.0025	0.0092	0.0072	0.0072
Options Market	0.4977	0.4022	0.3275	0.4097	0.3301	0.2748	0.5838	0.4919	0.3470	0.6418	0.5572	0.3725
Option Open Interest	20,485	1,592	91,664	35,461	4,932	123,690	4,468	585	15,095	1,741	272	5,406
Option Volume	3,558	145	22,587	5,458	309	28,751	663	52	3,114	299	41	1081
Option Notional Volume ($\times 10^6$)	44.86	0.37	690.01	76.82	1.44	909.63	0.96	0.08	4.88	0.14	0.02	0.56
Panel B: Specific informational trading days												
Stock Volume on EADs												
Total Share Volume ($\times 10^6$)	5.51	2.31	9.46	8.47	4.93	11.17	2.06	1.28	2.48	1.06	0.72	1.12
Off-exchange Share Volume ($\times 10^6$)	2.19	0.85	4.02	3.31	1.83	4.69	0.92	0.54	1.19	0.55	0.36	0.61
On-exchange Share Volume ($\times 10^6$)	3.32	1.44	5.49	5.16	3.09	6.53	1.14	0.74	1.31	0.51	0.36	0.51
Option Volume on EADs	10,846	2,781	24,661	17,743	8,084	30,434	1,986	1,065	2,955	498	315	542
Percentage Quoted Spread on EADs												
Stock Market	0.0034	0.0027	0.0029	0.0013	0.0011	0.0008	0.0039	0.0035	0.0018	0.0093	0.0086	0.0033
Options Market	0.4987	0.4291	0.2885	0.4108	0.3608	0.2159	0.5846	0.5371	0.2745	0.6450	0.6164	0.2301
Stock Volume on QWDs												
Total Share Volume ($\times 10^6$)	2.34	0.72	8.31	3.49	1.27	10.26	0.96	0.42	2.88	0.60	0.21	1.71
Off-exchange Share Volume ($\times 10^6$)	0.70	0.17	3.33	1.00	0.29	3.92	0.33	0.09	1.50	0.27	0.07	0.94
On-exchange Share Volume ($\times 10^6$)	1.64	0.54	5.15	2.49	0.97	6.52	0.64	0.32	1.43	0.33	0.14	0.80
Option Volume on QWDs	2,856	143	16,015	4,297	310	20,233	587	50	2,599	262	35	855
Percentage Quoted Spread on QWDs												
Stock Market	0.0025	0.0016	0.0031	0.0011	0.0009	0.0009	0.0031	0.0026	0.0020	0.0084	0.0067	0.0060
Options Market	0.4918	0.3925	0.3348	0.4063	0.3235	0.2823	0.5793	0.4825	0.3573	0.6486	0.5610	0.3751

A3 Market Shares for Various Exchanges

Figure A1 displays the time series of daily market shares for various exchanges from 2008 to 2022. We find the increasing popularity of off-exchange trading is coupled with a drastically decreasing volumes on NYSE, while NASDAQ and other exchanges trading has plateaued in the recent decade. Figure A2 divides NASDAQ stocks into two groups based on their listing exchange. Nasdaq was historically an off-exchange venue before it became an exchange in 2006 compared to NYSE which was registered as the main US exchange since 1792. In Daily TAQ database, NASDAQ uses two exchange IDs: T and Q. The "T" trades correspond to the trades on Tape A and B securities which include NYSE listed securities(Tape A) and NYSE Arca and NYSE Amex listed securities (Tape B). The "Q" trades correspond to the trade on Tape C securities which are NASDAQ listed securities. We note that the loss in market share for NASDAQ is concentrated on NYSE listed stocks. For Nasdaq listed stocks, the exchange is able to retain and even grow the market share in recent years suggesting that it has been able to retain it more innovative character from its off-exchange origins.

Figure A1. Time Series of Market Share for Different Exchanges

Off-exchange trades are identified by the exchange code "D", while trades with the exchange code "N" are labeled as NYSE trades. Trades on NASDAQ include codes of "T" and "Q", while all other trades are categorized as "Other". The sample period covers January 1, 2008, to December 31, 2022.

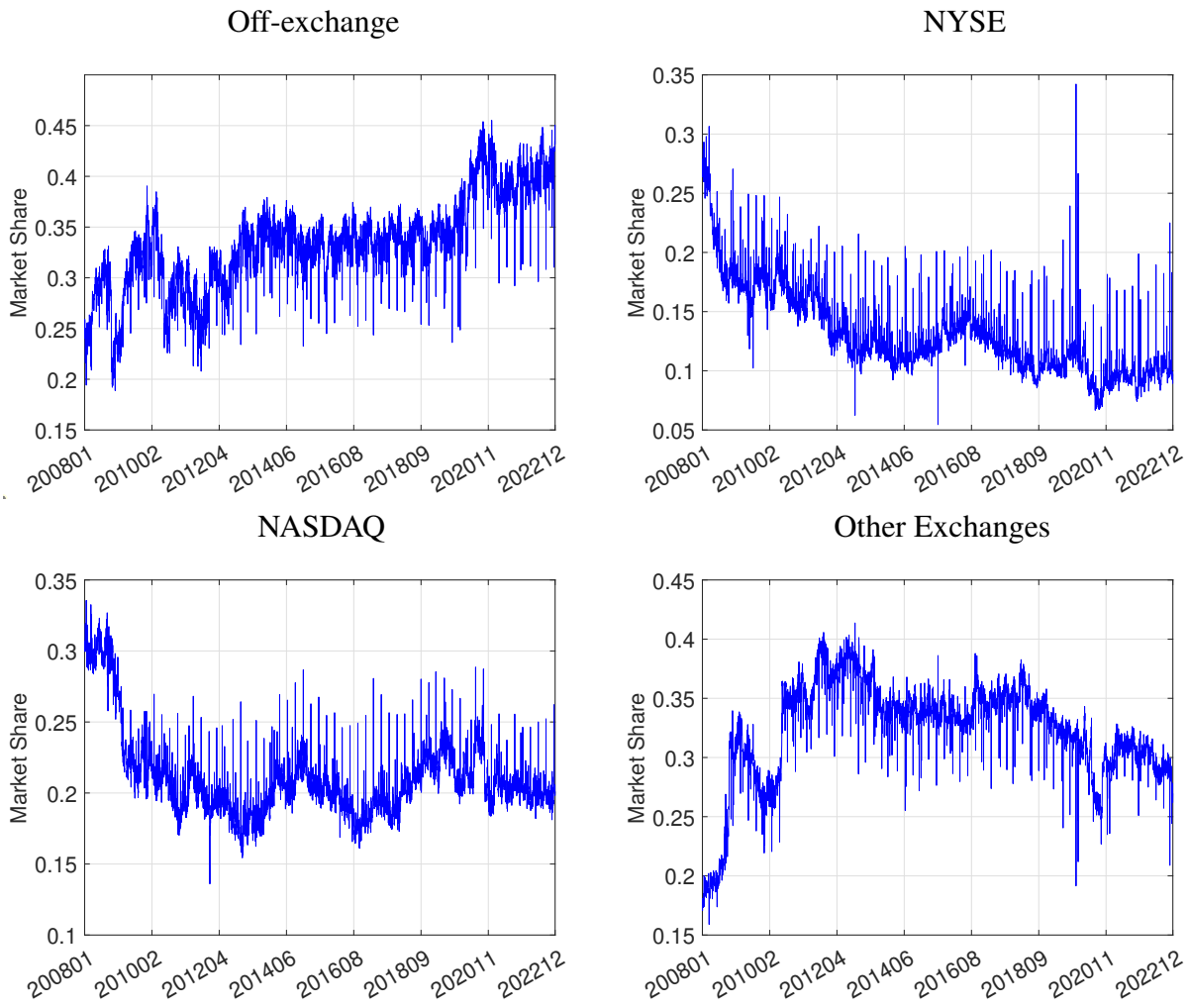
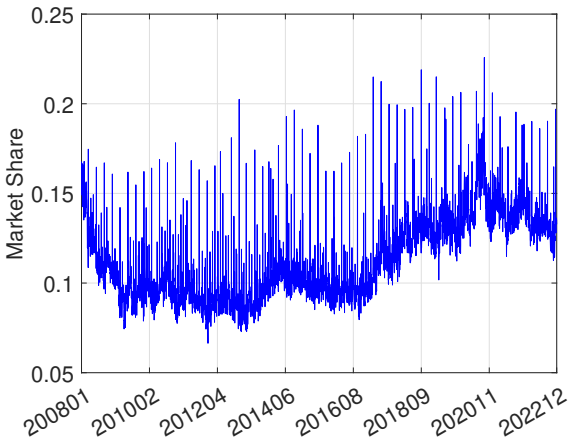


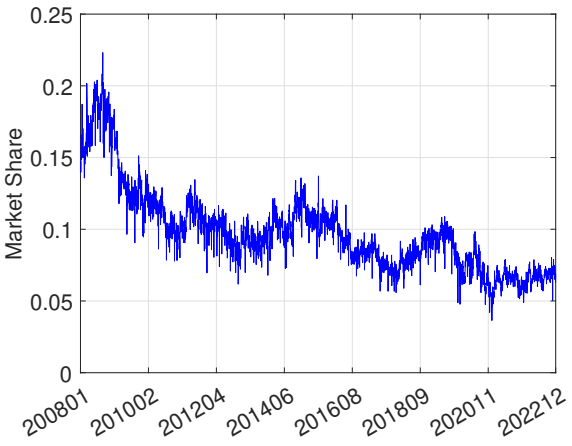
Figure A2. Time Series of Market Share for NASDAQ

Upward spikes patterns only appear on NASDAQ-listed securities in NASDAQ. Trades with the exchange code "Q" are NASDAQ trades on Tape C securities. Trades with the exchange code "T" are NASDAQ trades on Tape A and B securities. The sample period covers January 1, 2008, to December 31, 2022.

NASDAQ-Tape C



NASDAQ-Tape A,B



A4 Earning Days and Times

Table A3 shows the earnings release time and month for optionable and non-optionable stocks. First, we see that for optional stocks, regardless of the firm size, over 90% of firms release earnings outside the trading hours. Large firms have no preference between before and after trading hours but medium and small firms release earnings after hours. We have a similar pattern for the non-optionable stocks. Second, 45% large optionable firms release earnings in January, April, July and October. Remaining firms, optionable or not, release earnings in the second month of the quarter.

Table A3 Distribution of Earnings Announcement Days

This table displays the distribution of 137,075 earnings announcement days for optionable firms and 68,163 earnings announcement days for non-optionable stocks for 8,184 unique firms. Data period covers from January 1, 2008 to December 31, 2022.

	All	Large Firms		Medium Firms		Small Firms	
	Count	Count	Percent	Count	Percent	Count	Percent
Panel A: Optionable Stock							
Full Sample	137,075	72,618	52.98 %	50,688	36.98 %	13,769	10.04 %
Announcement Time							
0:00-9:30	52,011	31,345	43.16 %	16,265	32.09 %	4,401	31.96 %
9:30-16:00	14,692	7,979	10.99 %	5,318	10.13 %	1,395	10.13 %
16:00-23:59	70,372	33,294	45.85 %	29,105	57.42 %	7,973	57.91 %
Total	137,075	72,618	100.00 %	50,688	100.00 %	13,769	100.00 %
Announcement Months							
Jan/Apr/Jul/Oct	48,762	32,777	45.14 %	13,994	27.61 %	1,991	14.46 %
Feb/May/Aug/Nov	74,162	34,770	47.88 %	30,305	59.79 %	9,087	66.00 %
Mar/Jun/Sep/Dec	14,151	5,071	6.98 %	6,389	12.60 %	2,691	19.54 %
Total	137,075	72,618	100.00 %	50,688	100.00 %	13,769	100.00 %
Panel B: Non-Optionable Stock							
Full Sample	68,163	3,668	5.38 %	21,019	30.84 %	43,476	63.78 %
Announcement Time							
0:00-9:30	19,095	1,465	39.94 %	5,978	28.44 %	11,652	26.80 %
9:30-16:00	13,879	636	17.34 %	4,157	10.88 %	9,086	20.90 %
16:00-23:59	35,189	1,567	42.72 %	10,884	51.78 %	22,738	52.30 %
Total	68,163	3,668	100.00 %	21,019	100.00 %	43,476	100.00 %
Announcement Months							
Jan/Apr/Jul/Oct	21,171	1,364	37.19 %	8,421	40.06 %	11,386	26.19 %
Feb/May/Aug/Nov	36,605	2,030	55.34 %	10,114	48.12 %	24,461	56.26 %
Mar/Jun/Sep/Dec	10,387	274	7.47 %	2,484	11.82 %	7,629	17.55 %
Total	68,163	3,668	100.00 %	21,019	100.00 %	43,476	100.00 %

A5 Market Opening Auction

Table A4 illustrates a notable contrast between open trade volumes on quad-witching days and non quad-witching days. Specifically, it reveals that the open trade volume surges ten times higher during quad-witching days compared to non quad-witching days. Furthermore, when considering the average total volume, quad-witching days exhibit a trading increase of 1.06 billion shares in comparison to non quad-witching days, with approximately 66.22% of this increment attributable to the rise in open trade volume.

Table A4 Descriptive Statistics on Market Opening Auction Trade

This table displays the open trade statistics for quad-witching and non quad-witching days during the sample period. We focus on the trades executed during regular trading hours. Data period covers from January 1, 2008 to December 31, 2022.

	Non Quad-witching Days	Quad-witching Days
Open Trade Volume	71.65 millions	763.55 millions
Percentage of Total Volume	1.59%	13.71%
Off-exchange Volume	1.68 billions	1.73 billions
Total Volume	4.51 billions	5.57 billions

A6 Robustness Check

A6.1 Off-exchange trading right after an option trade

Here, we explore the mechanisms driving this shift to off-exchange activity by focusing on the sub-sample of stock trades that occur right after an option trade. Specifically, we regress the probability of off-exchange trading on stock spreads and option trade size and price to test whether informed traders chase low-cost stealth trading in off-exchange venues after spreads widen. One advantage of the intraday trade level data is that spread is directly observable before each trade, and we do not need to estimate it in 2 steps like we did for daily data analysis. The regression model for trade-level analysis is:

$$\begin{aligned}\text{Prob(Off-exchange}|\text{After Option)} = \alpha + \beta_1 \text{Spread} + \beta_2 log(\text{Option Price}) + \beta_3 log(\text{Option Size}) + \\ \beta_4 I_{\text{Expiration Date}} + \text{Controls} + \varepsilon\end{aligned} \tag{20}$$

The dependent variable Prob(Off-Exchange |After Option) takes the value of 1 for off-exchange trade and 0 for on-exchange trade. 'Spread' is observed before first stock trade following the option trade. $I_{\text{Expiration Date}}$ is an indicator variable that equals 1 if the trade date is the expiration date and 0 otherwise. The control variables include the log of the size and price of the stock trade preceding the option trades, the NBO and NBB depth of the quote prior to the option trades, the time gap between the last quote before the option trades and the option trades, an indicator variable for whether the stock trade preceding the option trade is executed off-exchange, the half-hour time interval and the day trend variable.

Table A5 reports the results of this regression. The first column uses dollar spread as the key explanatory variable. The statistically significant positive coefficient of 0.099 confirms our

hypothesis that lower trading costs drive informed traders to off-exchange trading after options trade widen stock spreads. We also note that the bigger the size of the option trade which is indicative of more intense information, the higher the probability of off-exchange trade. The negative sign on options price also supports the notion that informed traders, who typically choose low-priced out of money options, are more likely to choose off-exchange trading. We control for option expiration dates to control for the downward spikes in off-exchange trading related to settlement price determination on exchanges on these dates.

To understand the economic significance of this result, We show that a one standard deviation or 0.91 dollar spread (3.14% basis points for percentage spread) increase leads to $0.099 \times 0.91 = 9.01\%$ (286.76×3.14 bps=9.00%) increase in the probability of trading off-exchange. This confirms that a higher trading cost on exchange is an economically significant motivation to trade off-exchange. In addition, one standard deviation increase in the log of option size results in an increase of $0.043 \times 0.74 = 3.18\%$, while one standard deviation decrease in the log of option price leads to in increase in probability of trading off-exchange by $0.022 \times 1.73 = 3.81\%$. Alternative measures of trading costs in columns 2 to 4 namely, percentage spread, change in dollar spread, and change in percentage spread, all yield similar signs and significance for the coefficients and therefore confirm the findings stated above.

Table A5 Trade-Level Probit Regression

This table reports the probit regression on Amazon stock trades after an option trade

$$\text{Prob(Off-exchange}|\text{After Option)} = \alpha + \beta_1 \text{Spread} + \beta_2 log(\text{Option Size}) + \beta_3 log(\text{Option Price}) + \beta_4 I_{\text{Expiration Date}} + \text{Controls} + \varepsilon$$

where 'Spread' is observed before first stock trade following the option trades. We use four spread variables: dollar spread, percentage spread, change of dollar spread and change of percentage spread. The control variables include the log of the size and price of the previous stock trade, the NBO and NBB depth of the quote prior to the stock trades, the half-hour time interval variable and the day trend variable. The sample period covers from January 1, 2020, to December 31, 2020. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	Dollar Spread (1)	Percentage Spread (2)	ΔDollar Spread (3)	ΔPercentage Spread (4)
Spread	0.099***	286.76***	0.026***	101.52***
	(314.95)	(332.98)	(24.96)	(34.57)
log Option Size	0.043***	0.043***	0.042***	0.042***
	(121.56)	(121.81)	(119.75)	(119.80)
log Option Price	-0.022***	-0.022***	-0.019***	-0.019***
	(-140.21)	(-143.80)	(-126.46)	(-126.43)
$I_{\text{Expiration Date}}$	-0.044***	-0.044***	-0.049***	-0.049***
	(-71.13)	(-71.22)	(-80.33)	(-80.33)
Controls	Yes	Yes	Yes	Yes
N	24.67M	24.67M	24.67M	24.67M
Adj R^2	0.017	0.017	0.013	0.013

A6.2 Winsorized Spreads

Table A6 Off-exchange Activity Regressions using winsorized predicted spread

This table reports the regressions for optionable stocks. The first stage regression on stock quoted spread using full sample data

$$Spread_{i,t}^{Stock} = \alpha_i + \beta_1 \text{Spread}_{i,t-1}^{Option} + \beta_2 \text{Volatility}_{i,t-1}^{ATM} + \beta_3 \Delta\text{Open Interest}_{i,t-1} + \text{Trend}_t + \text{Controls}^*_{t-1} + \varepsilon_{i,t}.$$

Estimation results are reported in Table 4. We run the second stage regression on off-exchange measures

$$\text{Off-exchange}_{i,t} = \alpha_i + \beta_1 \widehat{\text{Spread}}_{i,t}^{Stock} + \beta_2 \text{Trend}_t + \text{Controls}_t + \varepsilon_{i,t}.$$

We retain all observations and substitute any negative predicted spreads with zero. The control variables (Controls_t) include SPY off-exchange market share, the market volatility index (VIX), log market capitalization, log price, log number of trades, intraday volatility, log dollar volume, and lag 1 dependent variable. The sample period covers from January 1, 2008, to December 31, 2022. The t-statistics are reported in parentheses. The symbols, ***, **, and * indicate significance at the 1%, 5% and 10% levels, respectively.

	Market Share	Information Share	Information Density
	(4)	(5)	(6)
$\widehat{Spread}_t^{Stock}$	3.23***	11.64***	37.02***
	(12.77)	(22.75)	(22.19)
Trend$\times 10^{-5}$	2.45***	0.33**	-1.03**
	(23.08)	(2.27)	(-1.99)
Controls	Yes	Yes	Yes
Firm Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
N	7.52M	6.70M	6.70M
Adj R^2	0.5129	0.2647	0.2076

A6.3 Earnings Announcement Days

Figure A3. Time-weighted Percentage Quoted Spread around Earnings Announcements
The sample period covers from January 1st, 2008, to December 31st, 2022.

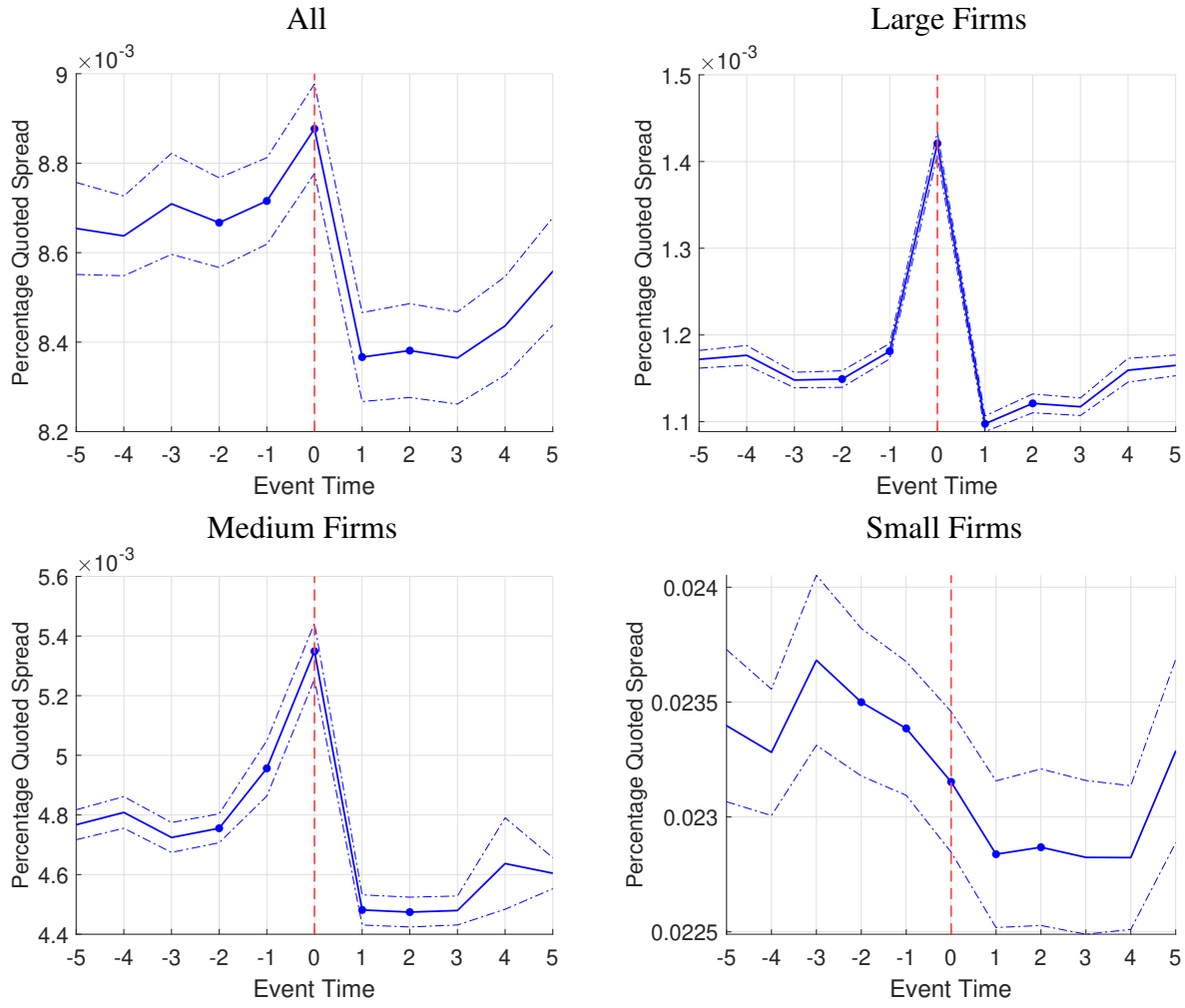
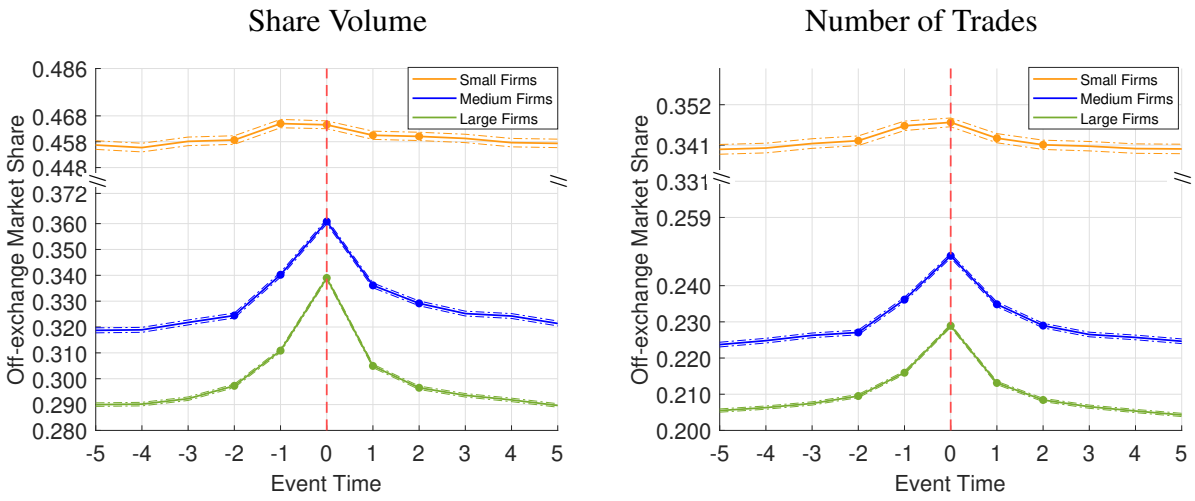


Figure A4. Off-Exchange Trading Market Shares Sorted by Firm Size Around Earnings Announcements

The sample period covers from January 1st, 2008, to December 31st, 2022.



A6.4 Robustness Check On Quad-Witching Days

To further confirm the increase trading volume on quad witching days mainly comes from on-exchanges, we run the following regression on different exchanges

$$\Delta\text{Volume}^{i}_{[t-1,t]} = \alpha_i + \beta_i I_t + \varepsilon_t \tag{21}$$

where $\Delta\text{Volume}^{i}_{[t-1,t]}$ is the daily change of the trading volume measures for exchange i from day $t-1$ to day t, and I_t is an indicator variable that takes the value of one if day t is the quad-witching day, and zero otherwise. We use three different volume measures: number of trades, share volume and dollar volume.

Table A7 Trading Volume on Quad-Witching Days

The table reports the daily change of the trading volume regression

$$\Delta\text{Volume}^{i}_{[t-1,t]} = \alpha_i + \beta_i I_t + \varepsilon_t.$$

Off-exchange trades are identity by the exchange code "D", while trades with the exchange code "N" are labeled as NYSE trades. Trades on NASDAQ include code "Q". Data period covers from January 1, 2008 to December 31, 2022.

	Number of Trades(unit: thousands)			Share Volume(unit: millions)			Dollar Volume(unit: billions)		
	Off-exchange	NYSE	NASDAQ	Off-exchange	NYSE	NASDAQ	Off-exchange	NYSE	NASDAQ
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	1.24	-2.41	-3.62	0.92	-6.42***	-7.01***	0.006	-0.32***	-0.33***
	(0.09)	(-0.39)	(-0.44)	(0.24)	(-3.41)	(-4.34)	(0.04)	(-4.20)	(-4.24)
I_t	41.94	132.59***	261.55***	-45.01	398.37***	443.40***	0.13	20.28***	21.16***
	(0.39)	(2.65)	(3.98)	(-1.50)	(26.42)	(34.33)	(0.13)	(32.89)	(33.78)
N	3776	3776	3776	3776	3776	3776	3776	3776	3776
Adj R^2	0.00004	0.00186	0.00418	0.0006	0.156	0.238	0.000004	0.223	0.232