Navigating the Murky World of Hidden Liquidity†

Robert P. Bartlett
Stanford Law School;
Stanford Graduate School of Business

rbartlett@law.stanford.edu

Maureen O'Hara
Johnson College of Business,
Cornell University

mo19@cornell.edu

October 2024
Revised, November 2024

Abstract

This paper investigates hidden liquidity on U.S. equity exchanges and how to find it. Despite the National Market System's (NMS) goal of transparent trading, we show that orders hidden from the NMS provide liquidity for 40% of the trading volume in our sample, rising to 75% of the dollar volume traded for high priced stocks. We demonstrate how price improvement on exchanges depends on interacting with hidden liquidity, especially with non-displayed orders. Leveraging big data and machine learning, we develop an algorithm that dynamically predicts where price-improving non-displayed orders are likely to appear, illustrating how AI-driven models can allow broker-dealers to more effectively meet their best execution obligations.

JEL classification: G14, G21, G23, G24
Keywords: hidden orders, Reg NMS, routing algorithms, execution quality, machine learning

† We thank London Stock Exchange Group (LSEG) for providing access to the proprietary data feeds of all U.S. stock exchanges, which provide the basis for the data used in this paper. We also thank Robert Battalio, David Easley, Thomas Ernst, Frank Portnoy, Ingrid Werner and seminar participants at the Microstructure Exchange and the Notre Dame Law School and University College London Law and Finance Symposium for very helpful comments. Additionally, some of the computing for this project was performed on the Sherlock cluster at Stanford University. We thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results.

Navigating the Murky World of Hidden Liquidity

1. Introduction

The U.S. National Market System (NMS) is based on a simple premise: the equity market should be transparent, allowing orders at any point in time to be routed to the exchange which has the best trading prices. While arguably true after the rules defining the NMS were implemented in 1975, much has changed over the intervening years, particularly with respect to transparency. Now, odd lot limit orders routinely exist between the NMS "best bid or offer", offering better trading prices than are reported in the national best bid or offer (NBBO) at the center of the NMS. These better quotes, however, are invisible to all but those able to purchase proprietary data feeds from the 15 current equity exchanges. Moreover, these better quotes are not "protected" by the NMS, meaning that orders need not be routed to them based on this hidden liquidity. Further complicating market transparency are non-displayed orders, which U.S. exchanges were permitted to offer starting in 2008. These orders, while losing execution priority to similarly-priced displayed orders, allow traders to post limit orders anywhere in the order book that are invisible to everyone in the market.[1] The upshot is that exchange-based trading is now far from transparent, with liquidity fragmenting into visible and invisible components. What hasn't changed in the market is the obligation of broker-dealers to get their customers best execution—but what does this even mean in a market where large amounts of liquidity at better prices are unobservable to market participants? And, given this, how do brokers-dealers satisfy the SEC's recent instruction to find this hidden liquidity across the 15 active equity markets by "identifying markets reasonably likely to provide most favorable prices"?[2]

To illustrate the importance of these issues, consider the following. Looking across all reported exchange trades for Google (GOOG) on January 5, 2023, we find that 52% of order executions are against non-displayed orders, 39% are filled by odd lot orders, and only slightly less than 10% of executions involved displayed liquidity. Looking across reported trade executions in Amazon (AMZN) from January 3, 2023 through April 14, 2023, we find that

[1] A third change is the rise of off-exchange trading, which FINRA data now places at approximately 45% of equity trading volume. Off-exchange trading has never been part of the National Market System so we do not address this important source of hidden liquidity in this paper. For more discussion of off-exchange equity trading see Bartlett and McCreary (2019) and https://www.cboe.com/insights/posts/off-exchange-trends-beyond-sub-dollar-trading/#:~:text=Off%2Dexchange%20U.S.%20equity%20volumes,billion%20shares%20in%20April%202023.

[2] See Securities and Exchange Commission 17 CFR Parts 240 and 242 [Release No. 34-96496; File No. S7-32-22] RIN 3235-AN24 Regulation Best Execution

exchange trades executed against displayed orders pay 100% of the half-spread unless they are for less than 100 shares. In contrast, similarly sized exchange trades that execute against non-displayed orders pay just 58% of the half-spread on average, and this ticks down to 56% of the half-spread for trades that execute against displayed odd lot orders. These two examples show that hidden liquidity is substantial and highly consequential for execution quality, but, of course, the real question is: how important is hidden liquidity for the market more generally?

In this paper we investigate the murky world of hidden liquidity. Understanding the scale of hidden liquidity is interesting in its own right, particularly in light of the many changes (technology, algorithmic trading, smart order routers and placers, high stock price levels, etc.) affecting equity market behavior. Equally important are the implications of this hidden liquidity for execution quality. Market venues are required to report metrics for "price improvement" or executions that take place at prices better than the posted NBBO.[3] As we demonstrate, however, there are really only two avenues for attaining such improvement in an exchange setting: odd lot orders and non-displayed orders inside the NBBO. Yet these are the order types that are hidden from NMS. And while brokers might find price-improving odd lot orders by subscribing to exchanges' proprietary data feeds, doing so will not reveal non-displayed orders.[4] This fact highlights a fundamental problem in the current, far-from-transparent National Market System: how do you go about finding this price-improving liquidity when it is, by definition, not displayed? A major contribution of our research is proposing a mechanism for doing so.

We begin by analyzing the scale of hidden liquidity in equity exchange markets. Using U.S. trade and order data maintained by the London Stock Exchange Group (LSEG), we exploit

---

[3] SEC Rule 605 mandates that market centers that trade National Market System securities to disclose monthly statistics on execution quality. As explained by FINRA "These reports include information about each market center's quality of executions on a stock-by-stock basis, including how market orders of various sizes are executed relative to the public quotes. These reports must also disclose information about effective spreads (the spreads actually paid by investors whose orders are routed to a particular market center). In addition, market centers must disclose the extent to which they provide executions at prices better than the public quotes to investors using limit orders." See https://www.finra.org/rules-guidance/guidance/sec-rule-605

[4] Additionally, the SEC has adopted changes to the definition of what constitutes a round lot to facilitate greater transparency in the NMS for orders of less than 100 shares. Under the new rules, a round lot order (which can form the basis of the NBBO) will depend on a stock's price. For stocks with prices at or below $250.00, a round lot stays at 100 shares; for stock prices greater than $250.00 and up to $1000.00, a round lot would be 40 shares; for stocks above $1000.00 up to $10,000.00, 10 shares; and for stocks above that level, 1 share. Using exchanges' proprietary data feeds, Bartlett, McCreary, and O'Hara (2023) use a counter-factual approach to investigate how these rule changes would have affected odd lot quoting between January and March 2021 and find that a substantial proportion of better-priced, non-round lot orders would remain hidden from the consolidated tape even with these new rules, requiring traders to purchase additional exchange data to see all displayed orders that are not round lots.

the fact that the LSEG trade data is constructed directly from exchanges' proprietary data products, which includes data not available through the conventional Trade and Quote (TAQ) data. Among other things, LSEG data includes the full order book of displayed quotes across all exchanges, allowing us to observe which intraday odd lot trades executed against displayed odd lot orders. And while we cannot directly observe non-displayed orders, the LSEG data identifies trades that involve these orders—thus giving us insight into executed non-displayed orders that is not available to researchers using TAQ data. We use these executed non-displayed orders along with the executed odd lot orders to proxy for the underlying hidden liquidity. Using a sample of stocks stratified by volume and price level, we determine the incidence of hidden liquidity for various stock sub-groups.[5]

We find a surprising complexity to hidden liquidity, with three results particularly striking. First, across our sample of $467 billion of trading volume, hidden orders provide liquidity for nearly 40% of it, and this fraction increases notably for stocks with higher-prices. For instance, for stocks priced between $1.00 and $5.00 per share, non-displayed orders provided liquidity for approximately 36% of the dollar value traded in an average stock-day-minute, while displayed odd lot orders provided liquidity for approximately 6% of dollars traded in such a minute. In contrast, for stocks priced at $200 per share and higher, these fractions increase to over 46% and 26%, resulting in approximately 73% of the dollars traded executing against orders that are hidden under Reg. NMS.

Second, hidden liquidity has critical implications for execution quality. We find that hidden orders are commonly placed as limit orders having better prices than an exchange's best displayed bid or offer. As a result, a marketable order crossing with such an aggressively-priced hidden order is likely to execute at a price that is superior to the NBBO. Indeed, we find that, depending on stock price, price improvement rates arising from executions against hidden liquidity ranged from 17% to as much as 72% of the dollar volume traded in an average stock-day-minute.

[5]As a regulatory matter, this definition of hidden liquidity focuses on those orders that are formally excluded from the definition of a "bid" or "offer" under Regulation NMS, thus allowing us to examine whether hidden liquidity compromises Reg. NMS' ambition of providing a transparent and equitable view of the best prices available for trading across U.S. exchanges.

Third, we delineate the mechanism whereby hidden liquidity facilitates price improvement relative to the NBBO. We decompose any price improvement for our sample of trades into the portion provided by non-displayed orders and the portion provided by displayed odd lot orders. For trades of 100 shares or larger, non-displayed orders provide all of the observed price improvement. For smaller trade sizes, price improvement arises from a combination of non-displayed orders and displayed odd lot orders. Quantifying the effect, we show that trades hitting non-displayed liquidity saved, on average, between 40% and 70% of the quoted half spread depending on the trade size. Trades interacting with displayed odd lot orders averaged savings of between 5% and 25% of the half spread depending on the trade size. We additionally demonstrate the important role that the price grid plays in determining these price improvement outcomes: In particular, a wider grid permits more opportunities for traders to submit hidden price-improving orders.

Finally, we turn to the fundamental problem posed earlier: Can we predict where in the market to find this hidden liquidity? Given the SEC's expectation that brokers will search for hidden liquidity, we assume brokers will use exchanges' proprietary data feeds to locate odd lot quotes; therefore, we focus on predicting non-displayed orders resting on an exchange at prices better than the SIP NBBO.[6] Our analysis compares the predictive ability of a model using standard statistical inference to that of a model based on machine learning methods. For each analysis, we use data from the first four months of our sample period as a "training period" and from the last two months as a "testing period". We use estimates from each model to generate predicted probabilities that a price-improving non-displayed quote (denoted PINQ) will be present on exchange *x* for stock *i* for each minute *t*. These probabilities give us the basis for a routing table to navigate to the non-displayed liquidity.[7]

---

[6] Indeed, public disclosures indicate that the major market makers already purchase the exchanges' proprietary data feeds to comply with their best execution duties, and these feeds provide information on where to find displayed odd lot quotes. See, e.g., Market Data Infrastructure, 85 Fed. Reg. 16,726, 16,752 n.284 (proposed Mar. 24, 2020) (codified at 17 C.F.R. pts. 240, 242, 249) (summarizing a Citadel representative's statement: "proprietary feeds are required for best execution"); Letter from Douglas A. Cifu, Chief Exec. Officer, Virtu Fin. Inc., to Brent J. Fields, Sec'y, U.S. Sec. & Exch. Comm'n 4 (Oct. 23, 2018), https://www.virtu.com/uploads/2019/02/2018.10.23-Virtu%E2%80%99s-Comment-Letter-Roundtable-on-Market-Data-and-Market-Access.pdf [https://perma.cc/8F82-UL6E] ("Simply put, Virtu could not fulfill its obligations to its myriad of retail customers and institutional clients without full depth of book market data feeds and robust exchange connectivity features that the SIP feeds alone do not offer.") Non-displayed odd-lot orders are included in our analysis as these are not visible on the proprietary feeds.

[7] Modern trade execution generally relies on execution management systems that employ algorithms to route orders to particular trading venues. Generally, these algorithms produce a routing table, or a sequence of which venues to

We present compelling evidence that our XGBoost machine learning model is able to "find" non-displayed price-improving liquidity in the market. We say that an exchange is likely to have non-displayed price-improving liquidity for stock i in minute $t+1$ if the model's predicted probability is greater than a threshold level of 50%. We find a predictive accuracy of 74.9%, and, of the 4.9 million stock exchange minutes having PINQs in our test data, the XGBoost model correctly identified 3,160,500 for a "recall" of 64.5%.[8] Using a more conventional statistical model, these metrics fall to 63.4% accuracy and a dismal 15.3% recall. Equally impressive, the XGBoost model performs strongly across all threshold levels, with an AUC of 89%.[9] We additionally use the Shapley Additive exPlanations (SHAP) method to explore which features of the model drive its strong predictive capacity. To our knowledge, this is the first paper to provide such a predictive model of non-displayed liquidity.

Our paper contributes to a large literature on hidden liquidity. Early literature (see Bacidore, Battalio, and Jennings (2002); Tuttle (2006)) looked at the effects of hidden depth at the quotes arising floor broker and market maker activities. Other literature focused on off-exchange trading and its effects on price efficiency (see, for example, Bloomfield and O'Hara (2000); Zhu (2014); Comerton-Forde and Putnins (2015); Foley and Putnins (2016); Hathaway et al (2017); Buti et al (2017); Topbas and Ye (2024)). Researchers have also looked at how the ability to hide orders in exchange settings affects traders' order strategies and market efficiency (see Moinas (2010); Boulatev and George (2013); Bloomfield, O'Hara and Saar (2015)). While most research finds that being able to hide orders either on or off exchanges induces changes in both informed and uninformed order strategies, the evidence on efficiency is less clear. Several recent papers look at hidden orders in the context of the 2016 SEC tick size pilot which widened tick sizes for small stocks. Edwards et al (2021) and Chung et al (2020) find that the resultant greater hidden liquidity reduced execution quality and decreased price efficiency. Pan, Van Ness,

---

try sequentially. A variety of factors can influence these algorithms such as maker/taker fees, volume discounts and specific trade algorithm features such as trade horizon. Battalio, Corwin, and Jennings (2016) argue that routing decisions appear to be motivated by rebates and not by best execution considerations.

[8] Accuracy measures the ratio of correct predictions / all predictions. Specifically, when the model predicts there will be a trade involving non-displayed price-improving liquidity in stock i on exchange x in the next minute, it is correct 74.9% of the time. Recall measures how accurate the model correctly identifies true positives from the actual positives in the data set. In an unbalanced sample as found here, recall is critical because it captures how well we find overall liquidity.

[9] The AUC of a model provides a measure of the power of the model to correctly classify across all classification thresholds.

and Van Ness (2024) additionally find that hidden liquidity in bond ETFs is significantly higher than in equity ETFs.

Our work here adds to this body of knowledge by providing insight into the current state of hidden liquidity on U.S. exchanges, a context that has been less studied in the literature due to data constraints. In earlier work (see Bartlett, McCreary, and O'Hara (2023)), we examined the displayed odd lot quotes that made up the "market within the market" and demonstrated the remarkable scale and informational importance of this inside market. Our analysis here shows a similar importance to the role played by non-displayed orders of any size, the other important dimension of hidden liquidity. We believe our results are particularly relevant for debates relating to current SEC efforts to modernize the National Market system (see related research by Ernst, Spatt and Sun (2024a); Sida, Ye and Zhang (2021)).

Our findings also emphasize the challenges that the growing role played by hidden liquidity on equity exchanges poses for best execution. Best execution in equity markets has long been problematic, in large part because as Macey and O'Hara (1997) argued the concept of best execution is multi-facetted and thus hard to define. As they note, even the courts have had challenges with what this concept actually means.[10] Research by Battalio, Corwin, and Jennings, (2016); Hung, Jorion, Lee and Schwartz (2023); Ernst, Malenko, Spatt, and Sun (2023); Dyhrberg, Shilko and Werner (2022); Battalio and Jennings (2023) investigates execution quality issues in retail and off-exchange trading.

Finally, our work contributes to the developing body of research applying machine learning techniques to address microstructure issues. Philip (2019) uses reinforcement learning to estimate the permanent price impact of a trade. Easley, Lopez de Prado, O'Hara and Zhang (2021) use random forest machine learning to investigate the predictive role of microstructure measures to predict futures market liquidity. Bartlett, McCreary, and O'Hara (2023) use XGBoost to examine the information content of odd-lot quotes inside the spread. Our analysis here shows how machine learning can be used to create a better process for finding the best price in equity markets, even when those prices are not displayed to the market.

[10] See, for example, the court's statement that "Fairness includes a duty of best execution, but the term itself eludes definition: despite its central nature it inhabits the texts or securities law without revealing its source". (/11 *re Merrill Lynch,* 911 F. Supp 754, 768 (1995))

2. Data and Sample Construction

Our intraday analyses rely on a set of proprietary data feeds from LSEG, the data vendor suppling the SEC with the microdata it uses for its Market Information Data Analytics System (MIDAS). The LSEG data capture all trade and displayable quote messages for all exchanges in the United States. Relative to the NYSE TAQ trade files, the LSEG data provide three distinct advantages for our study of hidden liquidity:

First, the LSEG data capture all displayed orders resting on an exchange regardless of size or price through the trading day, thus allowing insight into the size of orders resting across an exchange's order book at any point in time.

Second, while the proprietary data feeds do not include non-displayed limit orders, all exchanges other than IEX and NYSE American include an indicator in their data for whether the trade arose from the execution of a marketable order against a non-displayed limit order. LSEG normalizes these indicators by flagging each such trade with an "h" code.

Third, LSEG's trade data reflect the trades arising from the interaction of inbound market orders with each posted order on an exchange, thus reflecting the size and display characteristics of each individual limit order. This feature is particularly relevant for trades arising on the NYSE given that the NYSE consolidates trades arising from the same marketable order before distributing a trade report to the two Securities Information Processors (the "SIPs") responsible for disseminating the trade on the consolidated tape. Because TAQ data are based on SIP data, these individual limit orders are not reflected in the TAQ trade data for NYSE trades.

To illustrate, we present in Table 1 trades in the Class B shares of Greif Inc. (one of our sample securities described below) on January 3, 2023, between roughly 9:39:00 and 10:00:00. Columns (1) through (6) illustrate the date, time (stamped in nanoseconds), exchange, trade condition (if any), size and price of the trades as they appear in the TAQ trade data. Columns (7) through (12) include the time (stamped in microseconds), exchange, symbol, size and price as they appear in the LSEG data. In addition, columns (13) and (14) include information that is unique to the LSEG data, which are the direction of the market order if it interacted with a displayed limit order and an "h" flag for any trades that interacted with a non-displayed limited order.[11]

Overall, the TAQ data indicate that there were ten trades for a total of 437 shares traded, but the LSEG data indicate that there were thirteen trades for a total of 437 shares traded. The three extra trades (in bold) in the LSEG data arise from the fact that the buy market order at approximately 9:40:58 and the two sell orders at approximately 9:45:07 and 9:47:51 each executed against two limit orders, but each pair was consolidated into a single trade in the TAQ data.[12] In addition, note that the LSEG data indicate that two of the thirteen limit orders were not displayed at the time of the trade, including one that was consolidated into a single trade in the TAQ data.

LSEG's construction of its trade data based on each executed limit order allows us to track trades that interact with non-displayed orders, and we use these flagged trades to construct our sample of executed non-displayed limit orders for all exchanges other than IEX and NYSE American (neither IEX nor NYSE American flags such trades). For these latter two exchanges, we rely on the fact that both exchanges permit traders to post non-displayed midpoint orders;[13] therefore, we estimate the number of trades arising from non-displayed orders for these two exchanges by identifying trades executed at the midpoint of the NBBO reported by the SIPs (the "SIP NBBO") at the time of the trade based on LSEG data. In addition to the LSEG data, we also use MIDAS data for our descriptive statistics.

Our main sample encompasses trading days from January 3, 2023, to June 30, 2023. We use CRSP to construct a stratified sample of stocks that traded during this time period based on a stock's volume and market capitalization. In constructing the sample, stocks were first filtered to include only those with Standard Industrial Classification (SIC) codes outside the ranges 6000-6999 and 7000-7999 and having a share code (SHRCD) of 11. Stocks were additionally required to have data available for at least 208 trading days between September 1, 2022 and June 30, 2023. With these filters, stocks were grouped into quintiles separately for trading volume and

---

[11] The "s" flag indicates a market order that was marked as an intermarket sweep order (ISOs). ISOs are flagged as such within the TAQ data as an "F" indicator within the trade condition variable. The "I" indicator within the TAQ trade condition variable represents an odd lot trade.
[12] This can be seen by the fact that the trade times for each corresponding pair of trades in the LSEG data were identical to one another, and the total number of shares for each pair matches the size of each individual TAQ trade.
[13] See IEX, Order Types, https://www.iexexchange.io/products/order-types (describing non-displayed pegged midpoint orders); NYSE Group, Executed Order Type Usage, available at https://www.nyse.com/markets/nyse-american/trading-info#equities-order-types.

market capitalization over the 208 trading days, and strata were created by crossing the two quintiles. Finally, from each stratum, a random sample of 20 stocks was selected. As not all volume-by-capitalization strata contained 20 stocks, the final sample consists of 311 stocks across 21 strata.

We collect from LSEG all trade data as well as the price of the national best bid (NBB) and national best offer (NBO) at the time of each trade according to the SIPs. We use the NBB and NBO to classify midpoint trades and to calculate the amount of price improvement for each trade. For trades where LSEG did not provide trade direction, we also use the midpoint of the NBB and NBO to assign trade direction using the Lee-Ready method for all trades other than midpoint trades for which we do not assign a trade direction. Lastly, we drop all trades occurring before 09:35:00 and after 15:55:00, and any trades that reflect duplicate trades routed to other markets.[14] With these filters, our sample consists of 215,417,621 trades, of which 53,811,402 (25%) were flagged as "h".

Table 2 provides descriptive statistics. Consistent with our sample construction methodology, sample stocks ranged in their market values and trading volumes. The mean market capitalization of $3.3 billion was driven largely by the presence of several very large firms as reflected by the fact that the 25th percentile stock and the median stock had market capitalizations of just $87.9 million and $377 million. The distribution of dollar trading volume has similar characteristics, with the median of $4.2 million representing approximately one-seventh of the mean of $28 million. Seventy-five percent of stocks in the sample had an average stock price lower than $37, but the average of $38.5 also indicates several stocks that had stock prices considerably higher. The number of trades across stocks was slightly less skewed, with the average of 692,661 total trades being slightly double the median of 313,304. Of all trades during the sample period, on average roughly 28% executed against non-displayed orders and 66% reflected odd lot trades. These figures alone underscore the critical role of both non-displayed orders and odd lots in today's trading environment.

[14] In particular, comparison of the LSEG trade data with the TAQ trade data revealed that all four of the CBOE exchanges (EDGA, EDGX, BATS, BATSY) had approximately 10% more trades than were reported in TAQ for these venues. Moreover, all of these additional trades had a "Match ID Sale Condition" (which, when populated, is typically an alpha-numeric code) that began with "R", and dropping any trades having such a condition allowed the LSEG trade data to have a 1-for-1 match with the TAQ data. Correspondence with LSEG personnel confirmed that these trades reflected trades that were routed to away markets. Because these trades appear in the trade data of the away market, we exclude them to avoid double-counting.

Using the MIDAS data, we additionally examine how the level of non-displayed trading activity in our sample stocks compares across all stocks since 2019. As noted above, the MIDAS data is also constructed using the same LSEG data that we use, and the SEC provides daily tallies of the fraction of trades that execute against non-displayed orders on a stock-by-stock basis.[15] While these data cannot be used to conduct the intraday analysis we undertake, they nevertheless provide insight into the overall rate at which non-displayed orders appear in the trade data. In Figure 1, we plot the average daily fraction of trades marked as non-displayed for our sample securities as well as the rate for all other securities in the MIDAS data between January 3, 2019 and March 31, 2024. Both sample and non-sample stocks show a modest upward trend during this time period, with the mean rate of non-displayed trades for a stock hovering around 25% between 2019 and 2020 to around 27% since late 2022. The figure also suggests that sample stocks had a slightly lower non-displayed rate than non-sample stocks between late 2019 and most of 2020, but the difference disappears after this time. Indeed, after January 2021 the two groups become visually indistinguishable, suggesting that the overall non-displayed rate for our sample securities is no different than that for other securities during our sample period.

3. How (Un)Important Are Displayed Bids and Offers in Providing Liquidity?

Our roughly 215 million trade reports represent over $467 billion in trading volume. In this section, we explore who provided liquidity for these trades. Under Reg. NMS, the definition of a bid or offer is limited to round lot orders, though exchanges are permitted to aggregate odd lot orders for purposes of determining an exchange's best bid or offer (the exchange's "BBO").[16] One category of liquidity providers includes any trader who submits a limit order to buy or sell a security as either a displayed round lot or as a displayed odd lot that is aggregated with other orders to form a bid or offer of 100 shares or more on an exchange. A second category includes any trader that submits either an odd lot limit order that is not aggregated into a bid or offer of 100 shares or a non-displayed order at any price that is either a round lot or an odd lot. We refer

[15] These data are available to download from the SEC's website at https://www.sec.gov/data-research/market-structure-data.
[16] Updates to exchanges' BBOs are disseminated in real time on the consolidated tape. The NBBO is based on the best BBOs across exchanges.

to the first category of liquidity as “NMS Displayed” liquidity, and we refer to as the second as “NMS Hidden” liquidity.

It is useful to consider why a trader might provide NMS Hidden liquidity given its benefits and costs. First, consider an odd lot limit order that is displayed on an exchange but is nevertheless excluded from the consolidated tape. A trader submitting such an order may simply seek to post an order for a high-priced stock in a dollar amount that is insufficient to form a 100-share order, resulting in an odd lot order. Alternatively, a trader may submit a round lot limit order at an exchange’s BBO that is partially filled by a marketable order, causing the original BBO to “crumble” into an odd lot order priced better than the new BBO on the exchange.[17] Lastly, a trader may also use an odd lot limit order to minimize the price impact of placing an aggressively priced limit order: Because odd lot orders are not reported in the consolidated quote feed, the odd lot order will be less visible to the market as a whole.[18]

However, odd lot orders also come with a lower likelihood of being filled relative to displayed round lot orders priced at an exchange’s BBO. Under Rule 611 of Regulation NMS, only displayed round lot quotes can be “protected,” which means that no trading venue can execute a marketable order at a price that is worse than the price of the protected quote.[19] This “trade through” protection increases the likelihood that marketable orders will be routed to the best round lot quotes for execution, but odd lot orders can be “traded-through”.

Turning to non-displayed orders of any size, a trader submitting a non-displayed limit order gets the benefit of not signaling its trading interest, thus diminishing the likelihood that posting the order will cause prices to move against the trader. But like odd lots, these orders cannot be protected under Rule 611 as they are not displayed. Additionally, under exchanges’ time priority execution rules, these non-displayed orders are placed behind similarly priced displayed orders. Non-displayed orders are also hidden even from subscribers to an exchange’s proprietary data

---

[17] Similarly, a trader may post a less aggressive odd lot order that sits behind the BBO, but which is executed (or becomes the exchange’s best price) only if all orders at or better than the BBO are cancelled or filled, suggesting the trader is simply placing a small-sized passive order.

[18] It is also conceivable that a trader might post an aggressively priced odd lot order to avoid establishing a new NBB or NBO. Imagine, for example, a trader works a large sell order by placing an array of displayed and non-displayed orders, including non-displayed sell orders pegged to the midpoint of the NBBO. Placing an aggressively priced round lot order to sell that lowers the NBO will also cause the midpoint to decline, but placing the same order at 99 shares will preserve the existing NBO and the current midpoint price. Consistent with this theory, in unreported results we observe clustering of orders at 99 shares within the data, particularly for higher priced stocks.

[19] For more discussion of Rule 611 of the National Market System see https://www.sec.gov/spotlight/emsac/memo-rule-611-regulation-nms.pdf.

feed, making them invisible to traders and smart order routers searching for posted liquidity. To the extent a non-displayed order is filled, it will be because a marketable order happens to arrive on the exchange where the order is posted and either (a) the order is priced more aggressively than any displayed orders or (b) the marketable order is sufficiently large that it exhausts all displayed liquidity at superior or equivalent prices.

These rules regarding NMS Hidden liquidity suggest two general use cases for traders posting such orders. The first involves aggressive traders seeking to provide liquidity at prices better than the best displayed price on the exchange where the order is posted. While the quote has a lower likelihood of being filled for the reasons noted above, its aggressive price gives it execution priority relative to NMS Displayed orders when marketable orders arrive on the venue. The second use case involves placing an NMS Hidden order deeper in the order book. These orders will execute only when a large inbound marketable order arrives on the exchange. Given their low execution priority relative to NMS Displayed liquidity (and any NMS Hidden liquidity priced within the spread), these orders are truly passive insofar as their likelihood of execution depends on a large marketable order sweeping through one or more price levels of displayed orders.[20]

To assess the incidence of both kinds of NMS Hidden orders, we exploit the fact that the distribution of these orders at the top levels of an exchange's order book can often be inferred from our sample of trades. To see how this is possible, recall that the LSEG trade data for exchanges is based on individual limit orders that are filled by inbound market orders. Moreover, because LSEG records the time in microseconds reported by an exchange for each trade, it is possible to reconstruct the market order that triggered the corresponding trades in the LSEG data. For the same reason, it is also possible to reconstruct the top of the order book—displayed and non-displayed—at the time this market order arrived on the venue. Appendix A provides an example of how this can be done.

When we apply this technique to all trades in our sample, we find that our $467 billion of individual trade records can be linked to 151,425,017 marketable orders. Panel A of Table 3 explores how the 215,417,621 limit orders that provided liquidity for these marketable orders

[20] Five exchanges also offer Retail Liquidity Programs (RLPs) where liquidity providers can submit non-displayed orders at prices better than the NBBO for interaction with market orders from retail investors. Trades executed through an RLP are marked in our data as non-displayed and constitute only modest fraction of trading volume (see Ernst, Spatt and Sun 2024b).

were split across NMS Displayed orders and NMS Hidden orders.[21] Of the 215 million limit orders, roughly 18.7 million represented non-displayed round lot orders and 35.1 represented non-displayed odd lot orders. Thus, of the limit orders that provided liquidity for our sample of 151.4 million marketable orders, the Table shows that approximately 40% represented NMS Hidden liquidity and 60% represented NMS Displayed liquidity.

The table also shows that, to the extent a non-displayed order was a source of liquidity, it was almost always because the non-displayed order rested at the top of the order book: less than 1% of non-displayed odd lot or round lot orders provided liquidity to a marketable order at prices beyond the best price available on an exchange. In contrast, roughly 5% of displayed orders provided liquidity beyond the top of the book. Thus, to the extent a marketable order walked up or down the book, it was more likely to find liquidity with displayed orders than with non-displayed orders.

Panel B repeats the foregoing analysis, focusing instead on the dollar value of the limit orders that provided liquidity for our $467 billion of trading volume. Overall, the results are similar to Panel A. NMS Hidden orders provided liquidity for nearly 38% of the $468 billion of trading volume in our sample. Of this amount, $436 billion was traded at the top of exchanges' order books, of which over 39% was provided by NMS Hidden liquidity.

While Table 3 illustrates the overall importance of NMS Hidden liquidity across our sample of trades, the aggregation of trades over the sample period says little about what a trader placing a marketable order could expect to see throughout the trading day. The prevalence of NMS Hidden orders at the top of exchanges' order books also suggests we should find considerable heterogeneity in the importance of NMS Hidden orders in providing liquidity in low-price versus high-price stocks. The reason stems in part from Rule 612 of Reg. NMS, which requires orders to be priced in penny increments unless an order is for less than $1.00 per share. As stocks decline in price, the quoted spread can decline to a single penny, resulting in long queue lines at the best displayed bid and offer (Buti, Consonni, Rindi, Wen and Werner, 2015; Buti, Rindi, Wen and Werner, 2013; Kwan, Masulis & McInish, 2014, Chao, Yang, and Ye, 2017). As such, for low price stocks, there are fewer (if any) price points where a trader can use

[21] The number of displayed odd lots is underestimated as a trade that appears in LSEG as an odd lot could potentially be a partial fill of a round lot order. That is almost certainly the case for most lower priced stocks. So to be conservative, we only include as a "displayed odd lot" an odd lot trade that is better than the SIP NBBO, since there is no way for such a trade to come from a displayed round lot order.

NMS Hidden liquidity to place a more aggressive order than the exchange's BBO. In contrast, the wider quoted spreads associated with high price stocks create a larger price grid where traders can place aggressively priced NMS Hidden orders.

To investigate these issues, we examine the fraction of trading during a stock-day-minute that executes against NMS Hidden liquidity, and we do so by dividing each stock-day-minute into one of eight price-based groups. Observations in group 1 have average prices below $1.00 while group 8 stocks have average prices of $200.00 and over. Our sample is restricted to trading between 09:35 and 15:55; therefore, our estimates can be viewed as the expected fraction of liquidity provided by NMS Hidden liquidity in a sample stock during any of these 330 minutes when a trade occurred. Table 4 presents results for both the number of trades (Panel A) and the dollar volume of trading (Panel B). As the results are similar, we focus on the number of trades.

What is evident from Table 4 is that the stock price level has dramatic effects on both the incidence and composition of NMS Hidden liquidity. Whereas NMS Hidden orders provide 28.1% of the liquidity for the lowest price group, it provides 83.9% of the liquidity for the highest price group stocks. Perhaps a more telling statistic is that for stocks priced at or above $50, essentially 60% or more of liquidity provision arises from NMS Hidden liquidity. This ubiquity speaks to the dominant role that NMS Hidden liquidity plays for a large fraction of the stock market.

The results on the composition of NMS Hidden liquidity are equally intriguing. Non-displayed orders are positively related to stock price levels, ranging from 22% to 45% of all executed trades. But the relationship is non-linear, with stocks in price groups 2-7 all having non-displayed orders supplying approximately 35% of liquidity provision. Liquidity provided by odd lot orders is even more stock price dependent, ranging from 5.9% for the lowest priced stocks to 38.5% for the highest priced stocks. These results are essentially unchanged if we look at NMS Hidden liquidity at the top of the book or across all limit orders.

In general, the increasing relevance of NMS Hidden orders for higher priced stocks is consistent with the notion that traders use these orders to place aggressive orders between an exchange's BBO. In this regard, the low fraction of NMS Hidden orders for Group 1 stocks is somewhat surprising given the fact that orders for these stocks can be priced in sub-penny increments, thus creating more opportunities for traders to place aggressively priced NMS Hidden orders. One potential explanation may be the modest dollar volume of trades in Group 1 stocks shown in Panel B of Table 4. The fact that the average dollar volume of trades at the top of the book for Group 8 stocks is over \$100,000 during a minute while the average is just \$389 for Group 1 stocks may diminish a traders' concern that placing an aggressively priced, sub-penny order in a Group 1 stock will move prices against them.

Overall, our results in Table 4 underscore how the overall role of NMS Hidden liquidity shown in Table 3 masks considerable heterogeneity in how traders searching for liquidity can expect to encounter it during the trading day depending on a stock's price. Indeed, for a given market order, trades in high priced stocks are likely to interact primarily (perhaps exclusively) with NMS Hidden liquidity and trades in low price stocks are more likely to interact primarily (perhaps exclusively) with NMS Displayed liquidity. To illustrate, Figure 3 presents histograms of the distribution of stock-day-minutes based on the fraction of executed limit orders that are NMS Hidden for all Group 8 observations as well as for all Group 2 observations. For Group 8 stocks, for more than 40% of all stock-day-minutes, 100% of traded liquidity is provided by NMS Hidden orders while less than 1% interacted entirely with NMS Displayed orders. In contrast, for Group 2 stocks, NMS Displayed orders provided 100% of liquidity for roughly 25% of all stock-day-minutes.

Table 4 is also consistent with traders using NMS Hidden orders primarily to place aggressively priced orders that are superior to exchanges' BBOs. To be sure, the fact that so few market orders walk up or down an exchange's order book makes it difficult to assess the quantity of NMS Hidden liquidity resting deeper in the book. However, in Appendix B, we examine those market orders that do execute at multiple price levels, and we find that the likelihood of interacting with a non-displayed order declines as a market order walks up or down an exchange's order book.[22]

[22] More specifically, across the 151 million marketable orders in our sample, approximately 4 million walked up or down the order book. Among these orders, we can identify instances when a marketable order exhausted all liquidity (displayed and non-displayed) for a particular price level. For instance, a buy order that walked up five prices levels—as occurred in Figure 2A in the Appendix—exhausted all available liquidity at the first four levels. If we designate the top of the book as "level 0", we can compare the relative importance of displayed versus non-displayed liquidity for each price level *N* where the maximum price level for the market order is greater than *N*. Taking this approach, we find that the fraction of all orders that are non-displayed orders at price levels 1, 2, 3, 4 and 5 to be 22.4%, 11.4%, 10.44%, 9.6%, and 10.4%, respectively. We provide additional details regarding this methodology in Appendix B.

4. Hidden Liquidity, the NBBO and Price Improvement

That so much NMS Hidden liquidity represents an exchange's best available trading price suggests that much of this liquidity will be priced better than another central pillar of the National Market System: the NBBO. Because NMS Hidden orders are excluded from the definition of a bid or offer, they cannot establish the NBBO, which can only be set by the best displayed round lot orders across exchanges. In this section, we explore the implications of this observation for traders and the broker-dealers who are obligated to achieve best execution for them.

A. Are NMS Hidden Orders at the Top of the Book Priced Better Than the NBBO?

The NBBO is formed from the aggregation of exchanges' BBOs as they are updated in real time. As such, it should represent the "best" price to buy (the offer) or sell (the bid) for every U.S. listed equity security at each point in time during the trading day.

Despite this ambition, several factors prevent the NBBO from representing the actual best price to buy or sell a listed U.S. equity security. One fundamental problem is that the NBBO does not capture trading interest available within non-exchange venues such as in an alternative trading system (ATS) or by a retail market maker. During our sample period, 42% of the dollar volume of trading in sample securities occurred in these non-exchange venues. Moreover, to comply with both the trade-through rule under Rule 611 and a broker's best execution obligation, these trades are commonly executed at prices that are better than the NBBO. For instance, an ATS will typically allow a liquidity provider to submit orders that are pegged to the NBBO with a specified amount of penny-priced price improvement. Many ATS also allow orders pegged to the midpoint of the NBBO.[23] Likewise, a retail market maker that internalizes retail order flow will typically offer price improvement over the NBBO to comply with her best execution obligations.[24] However, in contrast to the rules pertaining to an ATS, retail market makers can offer price improvement with sub-penny prices, leading to the common practice among retail

---

[23] While the amount of price improvement must be incremented in pennies to comply with Rule 612 of Reg. NMS, an exception to this rule applies to orders priced at the midpoint of the NBBO. The largest ATS venues have long permitted traders to submit midpoint orders, thus enabling price improvement in these venues for highly liquid stocks with narrow spreads.

[24] As discussed in Ferrell (2001), the pressure to provide price improvement over the NBBO arose in large part due to the Third Circuit's decision in *Newton v. Merrill, Lynch, Pierce, Fenner & Smith, Inc.*, 135 F.3d 266 (3d Cir. 1998), where the Third Circuit found that a broker-dealer that automatically executed customer trades at the NBBO may not be in compliance with its best execution obligations. In addition, the regulations implementing Reg. NMS also created a perception within the industry that best execution may require a broker to seek out opportunities for customer price improvement (Bartlett & McCrary 2019).

market makers of filling retail orders with *de minimis* price improvement (Bartlett & McCrary, 2019).

Even for exchange trades, trading at the NBBO is often avoided in practice due to the existence of better-priced odd lot orders. As noted above, both the definition of the NBBO and the prohibition against trade-throughs have historically excluded odd lot quotes. Combined with the growing number of high-priced stocks with wide quoted spreads, the result has been the proliferation of odd lot quotes that sit within the NBBO and that can offer substantial amounts of liquidity (Bartlett, O'Hara and McCrary, 2023). Due to exchange's price-time priority rules, these odd lot quotes are also filled by inbound marketable orders ahead of quotes at the NBBO.

Aggressively priced non-displayed orders combine aspects of both "dark liquidity" within an ATS as well as odd lot quotes, providing yet another means for exchange trades to execute at prices that are superior to the NBBO. Like an ATS, exchanges permit liquidity providers to submit non-displayed orders that are either pegged to the midpoint of the NBBO or pegged to the exchange's best bid or offer with a specified amount of penny-incremented price improvement. And like a displayed odd lot order, non-displayed orders of any size operate within an exchange's price-time priority rules, so that when a marketable order arrives on an exchange, it first interacts with the best-priced displayed quote followed by the best-priced non-displayed quote. For a high-priced stock with a wide quoted spread at the NBBO, there might exist multiple better-priced NMS Hidden orders that can fill an inbound marketable order before the marketable order ever interacts with the exchange's best protected bid or offer.

In Table 5, we examine the extent to which the high fraction of liquidity provided by NMS Hidden orders shown in Table 4 translates into price improvement for trades in our sample. To do so, we return again to the stock-day-minute analysis used for Table 4. However, we focus here on NMS Hidden orders priced better than the NBBO, and we calculate for each stock-day-minute the fraction of all trades receiving price improvement due to NMS Hidden orders. For our measure of the NBBO, we use the SIP NBBO that prevailed at the moment of each trade according to LSEG data.

Overall, the results in Table 5 show similar patterns to those found in Table 4. The fraction of all trades receiving price improvement from both non-displayed limit orders and odd lot orders increases monotonically as stocks become more expensive. Moreover, comparing the figures in Table 5 with those in Table 4 reveals the extremely high likelihood that when a trade executes against either a displayed odd lot order or any non-displayed order, it will receive price improvement. In the case of odd lot orders, this is unsurprising given that, by construction, we identify displayed odd lot liquidity based on whether it is priced better than the NBBO. However, a non-displayed order at the top of an exchange's order book could be priced at the exchange's BBO, yet across all price groups, the fraction of trades that execute against any non-displayed order shown in Table 4 is roughly the same as the fraction of trades that execute against non-displayed orders priced better than the NBBO. This is especially true for higher priced stocks. For instance, non-displayed orders provided 45.4% of the liquidity for Group 8 stocks, but non-displayed orders with price improvement provided 44.3% of the liquidity for Group 8 stocks.

Table 5 also points to the different mechanisms by which non-displayed orders and odd lot orders can provide price improvement to trades relative to the NBBO. Ignoring the first row for stocks priced less than $1.00, the estimates in column (1) for non-displayed orders reveal a "U" shaped pattern based on a stock's price. Nearly one-third of trading in stocks priced between $1 and $5 executed against price-improving non-displayed orders, with the fraction declining moderately for stocks priced between $5 and $25, and rising again for stock prices above that. This U-shaped pattern reflects a similar trend found in non-exchange venues due to the narrowing of quoted spreads as stocks decline in price toward $1.00 per share. At these prices, quoted spreads can often collapse to $0.01, which is the smallest permissible spread owing to the combination of the prohibition against sub-penny quotes as well as the prohibition against locked or crossed markets.

These penny-constrained spreads produce long queue-lines for traders seeking to post orders on exchanges at the most competitive prices, leading traders to "queue jump" exchanges by offering to trade in an ATS at sub-penny prices (Buti, Consonni, Rindi, Wen and Werner, 2015; Buti, Rindi, Wen and Werner, 2013; Kwan, Masulis & McInish, 2014). Moreover, because of a combination of Rule 612 and an SEC enforcement policy regarding midpoint trades, these sub-penny trades are typically achieved through pegged midpoint orders (Bartlett & McCrary, 2019).

Consistent with the ATS trading results, our results in Table 5 show that midpoint trades constitute almost all of the Group 2 trades executing against non-displayed orders. Moreover, midpoint trades drop dramatically for stocks priced below $1.00. As the prohibition on sub-penny quotes does not apply to these securities, Group 1 stocks lack the long queue lines associated with stocks priced at $1.00 or higher (see Bartlett & McCrary, 2019), diminishing the need to queue-jump. In combination, these results suggest that traders use midpoint orders on both exchanges and non-exchange venues to engage in queue-jumping. Note, however, that the rate of midpoint trades declines monotonically as stocks increase in price. These figures indicate that the high rate of trades executing against price-improving NMS Hidden orders for higher priced stocks is therefore driven by traders exploiting the larger price grid created by wider quoted spreads of higher priced stocks.

Overall, these findings suggest that price-improving NMS Hidden liquidity reflects two primary forms of limit orders—those pegged to the midpoint and those priced at penny prices at the bid or offer—and that these two forms are a function of the size of the pricing grid available for placing permissible orders inside the NBBO. That is, where the grid is confined, traders will utilize non-displayed midpoint orders, and where the grid is spacious, traders will more commonly submit penny-priced bids or offers inside the NBBO. As in the ATS literature, we can test this proposition formally by means of examining the rate of these two forms of orders around the discontinuous change in the prohibition of sub-penny orders at $1.00 per share.

This is illustrated in Figure 3 where the blue dots represent the mean percent of shares traded at the midpoint during a stock-minute as a function of the volume-weighted average 2-decimal stock price for the minute. The red dots represent the same measure for shares traded against all other (non-midpoint) non-displayed orders. To highlight the discontinuous policy change regarding the use of sub-penny quotes for orders priced less than $1.00, we superimpose a vertical red line at $1.00. As shown in the figure, the rate of midpoint trades drops discontinuously as a stock's price falls below $1.00. In contrast, the rate of non-midpoint non-displayed trades increases discontinuously below this price. Using a local polynomial regression, we estimate a sharp regression discontinuity point estimate at the $1.00 cut-off of 0.23 (robust s.e.=0.0097) for midpoint trades, and -0.07 (robust s.e.=0.0123) for non-midpoint trades.

These results for non-displayed orders differ notably from the figures for displayed odd lots shown in Table 5. Displayed odd lot orders are subject to the prohibition on sub-penny quotes; therefore, except for Group 1 stocks, they are constrained by the size of the penny-denominated price grid within the NBBO spread. As a result, it is only for higher-priced stocks that better priced odd lot orders capture a sizeable portion of trades, with over 30% of trades in stocks priced above $50 per share executing against these orders.

B. Hidden Liquidity and Price Improvement

A clear implication of the prior analysis for traders seeking liquidity is the important role that NMS Hidden liquidity plays for price improvement. Long gone are the days when specialists on exchanges might offer price improvement on a discretionary basis. Rather, given regulations requiring quotes to be immediately and automatically accessible, whether a marketable order receives pricing better than an exchange's best displayed bid or offer is today dependent on the presence of NMS Hidden liquidity on the exchange.[25] The extent to which price improvement now depends on this form of liquidity has not been previously documented in the literature.

This phenomenon has special relevance for traders seeking to find price improvement relative to the NBBO, as will typically be the case for broker-dealers seeking to comply with their best execution obligations. Given the regulatory rules that limit trade executions at prices better than the NBBO, it is possible to decompose the price improvement that is available on exchanges into just two components: better priced odd lot quotes and non-displayed orders.

Figure 4 illustrates this decomposition. Using our full sample of 215 million trades, we focus on all trades for which the exchange provides the direction of the order, and we evaluate whether the trade received price improvement relative to the NBBO at the moment of the trade according to LSEG data. After excluding any trades flagged as ISOs, trades flagged as auction-priced, trades occurring during crossed or locked markets, and trades priced less than $1.00, we plot the fraction of trades that receive price improvement relative to the NBBO against trade sizes from 1 to 500 shares, and we do so separately for trades hitting displayed versus non-displayed orders.

As shown in the figure, trades hitting displayed orders received effectively no price improvement except for those trades interacting with displayed odd lot orders.[26] In contrast, the rate of price improvement for trades interacting with non-displayed orders was approximately 40% of trades for trade sizes between 200 and 500, and it was over 50% for trade sizes between

[25] Under Rule 611 of Regulation NMS, a displayed order will be protected against trade-throughs only if the quotation is immediately and automatically accessible.
[26] For trades greater than 99 shares, the mean rate of price improvement was 0.0003176, most likely due to modest variation in how venues calculate the NBBO relative to how it is observed by LSEG at the time of a trade.

100 and 200 shares. Moreover, rates were considerably higher than these levels for trade sizes at round lot intervals of 200, 300, 400 and 500, most likely due to the clustering of liquidity at these sizes. As with orders executing against displayed orders, there is also a notable and discontinuous increase in price improvement for trade sizes less than 100 shares. This last result is unsurprising given that non-displayed orders seeking execution priority at these odd lot sizes must compete for execution priority against any displayed odd lot orders resting inside the NBBO.[27]

In Figure 5, we repeat this exercise but focus on the amount of price improvement provided to orders of different sizes. To measure the amount of price improvement for a trade, we compare the execution price relative to the midpoint of the NBBO and calculate the percentage of the half-spread saved in the trade. For instance, a buy trade priced at the midpoint of the NBBO would save 100% of the half-spread, while a buy trade priced at the NBO would save 0% of the half-spread. As shown in the figure, the amount of the price improvement for non-displayed orders is substantial, particularly for trades at round lot intervals where average savings were roughly 60% of the half-spread. However, even aside from these trades, trades for less than 200 shares saved on average over 50% of the quoted half-spread. For non-displayed odd lot trades, savings for trades interacting with non-displayed odd lot quotes averaged 54%, which was more than four times the savings for trades interacting with displayed odd lot quotes.

Lastly, we additionally plot as hollow black circles the same measure of price improvement for all non-exchange trades in sample securities during the same time period based on TAQ data. Across all trade sizes, the amount of price improvement offered by non-displayed orders on exchanges is roughly the same magnitude as that offered by liquidity providers on non-exchange venues. The primary exception is for trades at round lot sizes where price improvement from non-displayed orders on exchanges is substantially higher than that provided on non-exchange venues. Likewise, trades for fewer than 100 shares also received slightly higher amounts of price improvement from non-displayed orders on exchanges than they did in non-exchange venues.

For broker-dealers seeking to get their customers best execution, Figures 4 and 5 have obvious implications. One is the value of having access to exchanges' proprietary data feeds

---

[27] Regression discontinuity estimates for the discontinuous change in price improvement rates at 100 shares are -0.06502 (robust s.e.=0.00219) for displayed orders and -0.07889 (robust s.e.=0.03387) for non-displayed orders, in each case based on Sharp RD estimates using local polynomial regression.

where a broker can observe better-priced odd lot quotes, as discussed in Bartlett, O'Hara and McCrary (2023). The other is the capacity to find non-displayed orders resting within the NBBO—a task to which we now turn.

5. Can We Predict Non-Displayed Liquidity? Finding the Orders You Cannot See

In this section we explore the ability to predict where non-displayed liquidity might exist across exchanges at prices that are inside the NBBO, which for convenience we refer to as price-improving non-displayed quotes (PINQs). The perspective we take is that of a market-maker or smart-order router (SOR) looking for non-displayed orders at prices inside the NBBO to supplement what they can observe on exchanges' proprietary data feeds; therefore, we ignore for this purpose odd lot quotes that can be viewed on these feeds. Fundamentally, hunting for PINQs is an exchange-specific enterprise. Therefore, we focus on the extent to which such a trader could use a conventional exchange routing table to rank exchanges by the likelihood of whether PINQs are present on the exchange.[28]

In conducting our analysis, our empirical approach compares the predictive capacity of routing tables based on conventional frequentist statistics compared to those based on machine learning (ML) methods. In both cases, we calibrate the predictive models using high-frequency, intra-day data to yield predictions that vary for each stock based on prevailing trade conditions. Thus, our analysis allows us to examine how automated, ML/AI-powered SORs can use real-time trading data to improve the capacity of traders to route trades to exchanges with price-improving, non-displayed liquidity.

For all analyses, we evaluate the predictive performance of a particular model by adopting the "firewall principle" (Mullainathan and Spiess, 2017) used in ML. Specifically, we split our

[28] A simple algorithmic strategy employed to find non-displayed midpoint liquidity is to send Immediate or Cancel (IOC) midpoint orders sequentially to each exchange. Given current latencies in equity trading, this strategy can potentially discover a non-displayed midpoint order. But, as we showed earlier, the rate of midpoint trades decreases monotonically with increases in the stock prices, a result largely explained by the larger price grid in which to place hidden orders. Thus, while useful, such an IOC strategy will be unable to find substantial amounts of non-displayed liquidity that is not at the midpoint. In addition, as noted by the SEC in its Proposed Order Competition Rule, pinging for midpoint liquidity at multiple venues can also increase the risk of information leakage, possibly resulting in some market participants canceling midpoint orders they have posted. See SEC Order Competition Rule 88 Federal Register 128, 210 n540. Perhaps for this reason, in an early effort to predict the location of non-displayed limit orders, the trading firm ITG Solutions emphasized the utility of a model that can predict price-improving non-displayed limit orders given that "[a]lgorithmic trading systems must either uniformly search across different market venues (at a great opportunity cost) or devise smarter ways to seek out available liquidity." (Bongiovanni, Borkovec and Sinclair, 2006)

dataset into a "training" dataset that we use to fit all models and then evaluate the predictive capacity of the model against a "test" (or hold out) dataset. To ensure our test and training data use the same sample securities, we split our dataset by time and use the first four months of data to train our models and the last two months of data to test them. This two-third/one-third split between training and test datasets ensures that the models are trained on a sufficiently large sample to capture the underlying patterns while retaining enough data to assess rigorously their out-of-sample predictive performance (Kohavi, 1995). This split results in a training dataset of 21,227,215 stock-exchange-minutes and a test dataset of 11,773,230 stock-exchange-minutes.

We begin by examining the overall rate at which trades interact with PINQs at each exchange within the training dataset. As a preliminary matter, there are at least two reasons to expect variation in the incidence of PINQs across exchanges. First, as discussed in Section 3, PINQs will only be executed if an in-bound marketable order happens to arrive at the venue where it is posted. As such, exchanges having a larger share of order flow should be more attractive to a liquidity provider posting a non-displayed order. Second, an exchange's pricing model may also influence the share of orders that are non-displayed rather than displayed. In general, exchanges either pay a rebate to liquidity providers and charge liquidity takers for each trade (maker-taker pricing) or pay a rebate to liquidity takers and charge liquidity providers for each trade (taker-maker or "inverted" pricing). In theory, the rebate offered to liquidity takers by inverted exchanges should create an incentive for traders to route their marketable orders to these venues, which would increase fill rates for—and thus, the incentive to use—PINQs. In addition, conventional wisdom among many traders posits that posting orders on inverted exchanges leaks information because the order implicitly conveys the trader's willingness to pay a fee (and forego a rebate) when posting to such a venue. While the empirical basis for this claim is contested (Bacidore, 2019), traders who subscribe to it may nevertheless view the use of non-displayed orders as a potential solution to this information leakage challenge on inverted exchanges.

Table 6 presents the rate of PINQs by exchange. Column (1) presents the mean percent of trades interacting with PINQs for each exchange across all stock-exchange-minutes for which at least one trade occurred, grouped separately according to the exchange's pricing model. To illustrate the trade frequency for each exchange, Column (2) presents the average number of trades across all stock-exchange-minutes for any observation where at least one trade occurred.

Across all exchanges, Nasdaq had the highest number of trades per stock-exchange-minute as well as the second highest rate of trades executing against PINQs, consistent with the notion that non-displayed orders will be more prominent on the most active exchanges. The top exchange by our outcome of interest was IEX, which ranked well behind many other exchanges in terms of the average number of trades executed. However, IEX has also engaged in extensive marketing to advertise its “rich midpoint liquidity,”[29] which may account for the high fraction of trades interacting with non-displayed orders on the venue.[30] Following Nasdaq and IEX, the next three highest exchanges by our outcome of interest were all taker-maker venues despite their having considerably less trading activity.

Given the distinction drawn above regarding non-displayed midpoint orders and non-midpoint orders, the adjacent columns explore how these averages change for stocks likely to differ in their respective incidences. Columns (3) and (4) limit the data to trades having a daily volume-weighted average price of more than $100, given that these stocks should have a wider pricing grid and therefore fewer midpoint quotes. Among these high-priced stocks, the percentage of trades that execute against PINQs increases markedly for all four taker-maker exchanges, ranging from approximately 37% of all trades on NSX to over 54% of all trades on EDGA. Trades executing against PINQs also increase on conventional exchanges, particularly at Nasdaq, Arca, BATS, NYSE and PSX where they constitute between 30% and 36% of all trades.

Finally, Columns (5) and (6) focus on trades in stocks with an average daily NBBO spread of a penny and a stock price of $1.00 or more, where PINQs are likely to be in the form of midpoint orders for reasons discussed previously. For these stocks, taker-maker exchanges no longer reveal significant levels of trades executing against PINQs. Instead, the top five exchanges by this measure are now IEX, Nasdaq, EDGX, NYSE American, and BATS. That non-displayed midpoint orders play such a small role on inverted exchanges would also suggest that the motivation for using non-displayed orders on these venues stems primarily from concerns about the information leakage arising from displayed limit orders.

---

[29] IEX Research, Maximizing the Midpoint on IEX Exchange, Oct 23, 2021.

[30] As noted previously, neither IEX nor NYSE American include a flag for non-displayed orders on their trade data; therefore, we used trades occurring at the midpoint of the SIP NBBO as a proxy for non-displayed orders on these two venues. The high rate of trades interacting with non-displayed orders on IEX shown in Table 6 would thus be consistent with IEX’s efforts to market the large amount of midpoint liquidity present on IEX.

These preliminary figures underscore the relevance of using both exchange-specific and stock-specific features for predicting the presence of PINQs on an exchange. Accordingly, we build our frequentist prediction model based on the following regression specification:

$$\begin{aligned} Y_{ixdt} = \beta_0 &+ \beta_1 \text{Price}_{ixdt-1} + \beta_2 \text{Spread}_{ixdt-1} + \beta_3 \text{Penny}_{ixdt-1} + \beta_4 \text{Return}_{ixdt-1} \\ &+ \beta_5 \text{Count}_{ixdt-1} + \beta_6 \text{Total\_Count}_{ixdt-1} + \beta_7 \text{Total\_ND}_{ixdt-1} \\ &+ \sum_{k=1}^{20} \beta_{k+7} Y_{ixd,t-k} + \gamma_i + \theta_x + \theta_x \times \text{Penny}_{ixdt} + \zeta_d + \eta_t + \epsilon_{ixdt} \end{aligned} \tag{1}$$

In the equation, Y_{ixt} is the percent of trades reported by LSEG that are flagged as non-displayed during each stock-exchange-minute. *Price* represents the natural log of the value-weighted average price from the prior trading minute, *Spread* represents the natural log of the quoted spread at the end of minute *t-1*, and *Penny* is an indicator for whether the quoted spread is less than 2 cents. All three of these variables reflect the close association between PINQs and the size of the pricing grid discussed in Section 4. In addition, we also include several variables related to a stock's overall trade frequency, which include: *Return,* the return on stock *i* over the prior minute; *Count,* the number of trades in stock *i* on exchange *x* during the prior minute; *Total_Count*, the aggregate number of trades in stock *i* on exchange *x* during the prior minute; *Total_ND*, the aggregate number of trades interacting with non-displayed orders in stock *i* on exchange *x* during the prior minute; and twenty lagged values of the Y_{ixdt} to capture serial correlation in the frequency of PINQs across stock-exchange-minute observations.[31] We additionally add γ_i to estimate the fixed effect for stock *i* and θ_x to estimate the fixed effect for exchange *x.* Lastly, to more precisely capture the relationship between the use of non-displayed midpoint orders on only some exchanges, we also include the interaction of exchange with *Penny*, and to capture any day and minute-of-the-day effects, we include fixed effects for date and minute-of-the-day.

We use estimates from this model to generate predicted probabilities that PINQs will be present on exchange *x* for stock *i* for each minute where a trade occurs in the test dataset. We

[31] Lagged values of *Y* are based on the prior stock-exchange-minutes and will, by construction, be zero if no trades occur during that stock-exchange-minute. Such a zero value, however, does not necessarily mean that no hidden orders rested on the exchange in the stock-exchange-minute given that we can observe hidden orders in the LSEG data only if a trade happens to occur on the exchange. In effect, this limitation of the data introduces a degree of measurement error into these lagged variables. Despite the potential for attenuation bias, the regression estimates we obtain for these lagged variables are particularly robust.

classify each stock-exchange-minute as likely to have PINQs if the predicted probability is greater than 50%. We then assess the model's performance based on its overall accuracy, precision, recall, and F1 score.

Regression estimates based on the training dataset are in Table 7. The table presents two sets of estimates to illustrate the significant serial correlation for our outcome of interest. Column (1) first presents the estimates from using a specification that omits the twenty lagged values of the outcome variable, while column (2) provides estimates for the full specification. Given the central importance of identifying exchanges with PINQs, both columns retain the coefficient estimates for the exchange fixed effects, which in column (1) again highlight the higher likelihood of PINQs on IEX, Nasdaq, and BATS, along with the inverted exchanges BATSY, BX, and EDGA. However, introducing the lagged values in column (2) causes the coefficients for some of the most active exchanges to lose significance (BATS, NASDAQ, NYSE), suggesting their exchange fixed effects can be captured by looking at recent trading activity on these exchanges. Relative to column (1), the R-squared for column (2) also increases from 18.0% to 23.1%. In both columns, the interaction of *Penny* and stock exchange captures the trend shown in Table 6 with regard to the lack of PINQs on inverted exchanges for penny-constrained spreads.

Of the other variables, the negative coefficient of -.067 on *Price* and the positive coefficient of roughly 0.09 on *Spreads* in column (2) highlight how the general association between stock price and PINQs is largely driven by the larger spreads associated with higher-priced stocks. PINQs also appear to be positively associated with stock returns over the prior minute. In terms of trading activity, the incidence of PINQs in a stock-exchange-minute is negatively associated with prior trading in the past minute for that stock-exchange pair, but it is positively associated with the overall level of trades for a stock that interact with non-displayed orders across all exchanges in the prior minute.

In Table 8, we present performance results for using these estimates to predict stock-exchange-minutes with PINQs in the test dataset. Overall, using estimates from column (1) of Table 7 to predict whether there is more than a 50% chance that a stock-exchange-minute will reveal PINQs in the test dataset, we correctly classify approximately 60% of the observations as revealing or not-revealing PINQs. Moreover, while the model's precision (the fraction of positive classifications that were in fact positive) was over 61%, the model's recall (the fraction of true positive observations that were classified as positive) was just 7.6%. More concretely, this poor recall is reflected in the fact that of the nearly 4.9 million stock-exchange-minutes revealing PINQs in the test dataset, the model classified just 370,826 as positive. This result drives the model's F1 score—which combines both precision and recall—to be just 13.6%.

Column (2) of Table 8 reveals that adding the lagged variables results in a notable improvement in predictive performance, with accuracy and precision both increasing to approximately 63.5% and 81.4%, respectively. However, recall remains low, with the model classifying as positive observations of PINQs just 15.3% of all stock-exchange-minutes where trades in fact executed against PINQs. Despite the improvement in precision, the model's F1 score remains just 25.8% due to this poor recall.

We next examine the performance results for Equation (1) (including lags) when calibrated using a machine learning (ML) methodology. Relative to regression models, ML algorithms can often have stronger predictive properties owing to their ability to fit complex and flexible functional forms to data in a way that uncovers relationships between a set of predictors and outcomes that are not specified in advance. Additionally, ML estimation utilizes regularization techniques that avoid over-fitting to maximize out-of-sample predictive performance. To evaluate the utility of relying on an ML approach to improve on the statistics shown in column (2) of Table 8, we employ XGBoost, a gradient boosting algorithm that is closely related to the more common random forest algorithm but contains several features that lead to modest improvements in a number of predictive tasks (Chen and Guestrin 2016). Within the finance literature, Bartlett, McCrary, and O'Hara (2023) illustrate how to deploy an XGBoost model using microstructure data and provide additional information on this particular ML model. We use it here to assess the ability to predict the exchanges where one is likely to find PINQs.[32]

Performance statistics for using XGBoost to estimate Equation (1) are in column (3) of Table 8. Relative to column (2), using XGBoost results in a notable gain in the overall predictive performance of the model. Accuracy increases from 63.58% to 74.88%, and while precision declines from 81.37% to 71.83%, recall increases dramatically from 15.3% to 64.45%. This

[32] XGBoost contains a set of hyperparameters Θ that jointly determine model representation and optimization. In particular, a number of these hyperparameters affect the degree of regularization imposed during model fitting. To select the best set of hyperparameters $\Theta\star$, we perform grid search across a large range of possible combinations of hyperparameters using 5-fold cross-validation. We select the hyperparameter set that maximizes the receiver operating characteristic during this process.

notable increase in recall results a much-improved F1 score of roughly 68%. More generally, while we used a 50% threshold for predicting whether PINQs appeared in a stock-exchange-minute, the model performs strongly across any threshold. Specifically, in Figure 6, we assess the receiver operator characteristic (ROC) area under the curve (AUC), which evaluates the true positive rate and the false positive rate of a model using different probability thresholds for a positive classification. The AUC of a model thus provides a measure of the power of the model to correctly classify across all classification thresholds. As shown in Figure 6, estimating Equation (1) using XGBoost results in an AUC of 89%.

As in Table 8, we additionally explore which features of the model might account for its enhanced predictive performance. While ML-methods lack the capacity to produce point estimates for each feature, the SHapley Additive exPlanations (SHAP) method (Lundberg and Lee, 2017) provides a means to explore which of the 352 features used in our XGBoost model contribute the most to estimating the predicted probability of PINQs for a given stock-exchange-minute observation. As discussed in Bartlett, McCreary, and O'Hara (2023), the SHAP method utilizes principles of game theory to estimate the marginal contribution of each feature in explaining an individual predicted outcome, relative to the expected probability within the data. For a classification model such as ours, a feature's SHAP value is estimated for each individual observation as the change in the predicted probability of a positive classification due to the addition of that feature, averaged across all possible combinations of the feature values for that observation.[33]

Figure 7 presents a "beeswarm" summary plot of the SHAP values for the top 20 features across all 21,227,215 stock-exchange-minutes in the training data. Positive SHAP values indicate a positive prediction of PINQs, and the features on the y-axis are ordered top to bottom according to their importance. The color of each point represents the high (red) to low (blue) values of the feature, thus providing insight into how the marginal contribution of a feature

[33] For a model with n features, calculating exact SHAP values theoretically requires evaluating 2^{n-1} feature combinations for each observation, leading to an infeasible 2^{351} combinations per observation in the case of a model with 352 features. Given the impracticality of computing SHAP values directly for an XGBoost model with such high dimensionality, we used the auto-shap library in Python, which leverages Tree SHAP for tree-based models such as XGBoost. Tree SHAP is an efficient approximation technique that reduces the exponential complexity by exploiting the model's tree structure, allowing for polynomial-time computation of SHAP values. Additionally, auto-shap further enhances computational efficiency by supporting multiprocessing, which allows SHAP values to be calculated in parallel across available CPUs (64 in our case). With over 22 million observations, deploying this approach on a high-memory (>1GB RAM) server cluster permits SHAP value estimation for all 352 features in roughly 5 hours.

changes according to its value. Across all features in the model, the natural log of the quoted spread has the highest overall contribution to the estimated predictions, with wider (narrower) spreads leading to higher (lower) predictions that PINQs appear in a given stock-exchange-minute. This result is largely consistent with the positive, significant coefficient on spreads in Table 7. Likewise, the SHAP values for the exchanges—Nasdaq (exchange_INET), IEX (exchange_IEXDEEP), and MEMX (exchange_MEMX)—are also consistent with Table 7, though it is notable that only these three appear within the top 20 features despite other exchanges (e.g., BATSY and BX) having higher point estimates in Column (2) of Table 7. Additionally, twelve of the top twenty features represent lagged values of the target of interest, underscoring the serial correlation of PINQs within the dataset.

Somewhat surprisingly, the minute of the day ("t")—which was included as a fixed effect in Table 8—also appears as a top 20 feature. As shown in Figure 8A, visual examination of the fraction of trades that are non-displayed across trading minutes reveals a notable increase during both the opening and closing minutes of the day, which accounts for the fact that both early (morning) and late (afternoon) minutes are associated with higher SHAP values in the beeswarm plot. As overall trading volume is higher at the opening and closing of the trading day (see Figure 8B), this finding also suggests traders may turn to non-displayed orders when volume is expected to be higher, which would increase the likelihood that a non-displayed order is filled.

Finally, in Figure 9, we provide an example of how using an ML-powered SOR with real-time training data can enhance the likelihood of identifying exchanges with PINQs. The graph illustrates the predicted probabilities of finding PINQs for the sample stock Zurn Elkay Water Solutions Corp. (ZWS) at 15:50 on June 27, 2023, across the 13 exchanges where trades in ZWS occurred during that minute. The height of each bar in the graph reflects probabilities from one of two models based on Equation (1): the full regression model (including lags) (orange) and with the same model estimated using XGBoost (green). Exchanges marked with an asterisk (*) indicate exchanges where trades in ZWS in fact executed against PINQs.

The average stock price of ZWS during our sample period was approximately $25, with an average quoted spread of just over 1 cent. Looking first at the regression predictions, the figure reveals that the three most likely exchanges to have PINQs were EDGX, IEX, and Nasdaq, and trades in ZWS did in fact execute against PINQs on these exchanges. However, for none of the exchanges were the predicted probabilities greater than 50%. As such, the model provides little confidence that routing trades to these three exchanges would create price improvement due to the presence of PINQs.

In contrast, estimates for the XGBoost model are higher than 50% for six stock exchanges, all of which had trades execute against PINQs during the stock-exchange-minute. Moreover, the model also ranked probabilities for Nasdaq and EDGX higher than for IEX, despite the high likelihood of PINQs on IEX for a stock such as ZWS shown in Table 6, while also (accurately) predicting PINQs on EDGA despite the low rate of PINQs on EDGA when spreads are penny constrained. Overall, recall for this observation was 100%, while it was 0% for the regression model.

To be sure, not all stock-exchange-minutes in the test data saw similarly strong performance from the full XGBoost model. Nonetheless, the figure illustrates the implications of the substantially stronger recall available with the ML-model and, more generally, the potential for using ML-powered SORs to exploit real-time trading data when searching for PINQs to discharge a broker's duty to find the best price to trade in the market.

6. Conclusion

The National Market System may aspire to provide a transparent equity market, but our results show the reality is far from transparent. Hidden orders provided liquidity for almost 40% of all exchange trading volume in our sample, with high-priced stock liquidity particularly in the dark: on average, approximately 75% of the dollars traded in these stocks executed against orders hidden under Reg. NMS. Because hidden liquidity is so ubiquitous, it is not surprising that the NBBO is far from representing the "best" trading prices available in the market. Indeed, we found that depending on stock price, from 38% to 72% of the dollar volume executed in an average stock-day-minute executed against hidden liquidity at prices better than the NBBO for stocks priced higher than $1.00 per share. Receiving "price improvement" in this world is thus almost an oxymoron: there are better prices already out there, so getting a better price than the posted quotes is more the norm than the exception.

The challenge, of course, is how to find this hidden liquidity. While traders can and do turn to exchange's proprietary data feeds for displayed odd lot quotes, incurring this expense still leaves one without a guide to non-displayed orders that we show are so important for delivering price-improvement. We proposed a mechanism to find these price-improving non-displayed orders using machine learning and inputs from the LSEG data set. In particular, using the LSEG data we showed how to find transactions that involved executed non-displayed orders that only appear in the exchanges' propriety data sets. This allowed us to estimate the incidence of non-displayed (executed) liquidity across exchanges. Combining this data with stock-minute trading data, we developed a dynamic algorithm for predicting the probability that a non-displayed price-improving order will be present for stock *i* on exchange *x* during minute *t*. Our results are impressive, suggesting that using AI-powered trading systems and SORs may be one solution to shed light on non-displayed stock market liquidity.

We stress, however, that while our machine-learning based approach should be a valuable tool for broker/dealers seeking to achieve their best execution obligations, it is not a total solution to the transparency problem. Like everyone else in the market, we cannot see the actual non-displayed orders, and so we use the executed non-displayed orders as a proxy for this form of hidden liquidity. As such, designing a predictive model based on the non-displayed orders that are executed will necessarily be less comprehensive than a model trained on data for all non-displayed orders, regardless of whether they are executed.

Providing true transparency will thus require changes to the Reg NMS structure. Already, the SEC has proposed several changes, many of which may enhance transparency and price discovery. These include changes to the minimum tick size, moving from a standard 1 cent tick to a variable tick taking on values of .001, .002, .005. and .01 depending on a stock's time weighted average spread.[34] This change will allow for sub-penny price executions and so should reduce the spreads paid by investors. The SEC has also proposed changes to the definition of a round lot (currently 100 shares), reducing the size depending on the stock price. This will help transparency by increasing the orders that are visible to the market (and included in the NBBO). A third proposed change is greater disclosure of odd lot orders to the market. These proposed changes, in our view, are steps toward achieving the transparency goal at the heart of Reg. NMS, but non-displayed orders will remain. One possible solution would be to require exchanges to retain and disclose on a historical basis non-disclosed orders just as they do for displayed orders, which could enable more precise predictive models. However, absent additional reform in this domain, the quest will remain to find all the orders you cannot see.

[34] These proposed changes are discussed in https://www.federalregister.gov/documents/2022/12/29/2022-27616/regulation-nms-minimum-pricing-increments-access-fees-and-transparency-of-better-priced-orders

References

Bacidore, Jeff, 2019, Do Inverted Venues Leak Information, *Traders Magazine*, December 2019.

Bacidore, J, R. Battalio, and R. Jennings, 2002, Depth Improvement and Adjusted Price Improvement on the NYSE, *Journal of Financial Markets* 5, 169-195.

Bartlett, McCrary, and O'Hara, 2021, The Market Inside the Market: Odd-lot Quotes, *Review of Financial Studies*, forthcoming.

Bartlett, R. and J. McCrary, 2019, Dark Trading at the Midpoint: Does SEC Enforcement Policy Encourage Direct Feed Arbitrage, *Journal of Law, Finance, and Accounting* 4, 291-342.

Battalio, R, S. Corwin, and R. Jennings, 2016, Can Brokers Have it all? On the Relation between Make-Take Fees and Limit Order Execution Quality, *Journal of Finance* 71, 2193-2237.

Battalio, R. and R. Jennings, 2023, Wholesaler Execution Quality Available at SSRN: https://ssrn.com/abstract=4304124

Bloomfield, R. and M. O'Hara, 2000, Can Transparent Markets Survive? *Journal of Financial Economics* 55, 425-459.

Bloomfield, R., M. O'Hara and G. Saar ,2015, Hidden Liquidity: Some New Light on Dark Trading, *Journal of Finance* 70, 2227-2274.

Bongiovanni, S., M. Borkovec, and R. Sinclair, 2006, Let's Play Hide-and-Seek: The Location and Size of Undisclosed Limit Order Volume, *The Journal of Trading* 1, 34-46.

Boulatev, A. and T. George, 2013, Hidden and displayed liquidity in securities markets with informed liquidity provision, *Review of Financial Studies* 26, 2096-2137.

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

Buti, Sabrina, Francesco Consonni, Barbara Rindi, Yuanji Wen and Ingrid Werner, 2015, Sub-Penny and Queue-Jumping, Working paper. The Ohio State University.

Buti, Sabrina, Barbara Rindi, Yuanji Wen and Ingrid Werner, 2013, Tick Size Regulation and Sub-Penny Trading, Working paper. The Ohio State University.

Chao, Y., C. Yao, and M. Ye, 2017, Discrete Pricing and Market Fragmentation: A Tale of Two-Sided Markets, with Yong Chao and Chen Yao, *American Economic Review: Papers and Proceedings* 107, 196–199.

Chung, K., A. Lee and D. Rosch, 2020, Tick size, liquidity for small and large orders, and price informativeness: evidence from the tick size pilot program, *Journal of Financial Economics* 136 9, 879-899.

Comerton-Forde, C. and T. Putnins, 2015, Dark Trading and Price Discovery, *Journal of Financial Economics* 118, 70-92.

Dyhrberg, A., A. Shkilko and I. Werner, 2023, The Retail Execution Quality Landscape, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4313095

Easley, D., M. Lopez de Prado, M. O'Hara and Z. Zhang, 2021, Microstructure in the Machine Age, *Review of Financial Studies* 34, 3316-3363.

Edwards, A., P. Hughes, J. Ritter, P. Vegella, and H. Zang, 2021, The Effect of Hidden Liquidity: Evidence from an Exogenous Shock, Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3766512

Ernst, T., A. Malenko, C. Spatt, and J. Sun, 2023, What does best execution look like? Working paper, University of Maryland.

Ernst, T., C. Spatt and J. Sun, 2024a, Would Order-by-Order Auctions be Competitive? *Journal of Finance*, forthcoming.

Ernst, T., C. Spatt and J. Sun, 2024b, Why Did Retail Liquidity Programs Fail? Working paper.

Ernst, T. and C. Spatt, 2023, Payment for Order Flow and the Retail Trading Experience, working paper.

Foley, S and T. Putnins, 2016, Should we be afraid of the dark? Dark trading and markets quality, *Journal of Financial Economics* 122, 456-481.

Hatheway, F., Kwan, A. and H. Zheng, 2017, An Empirical Analysis of Market Segmentation on U.S. Equity Markets, *Journal of Financial and Quantitative Analysis* 52, 2399-2427.

Hung, X., P. Jorion, M. Lee and C. Schwartz, 2023, Who is Minding the Store? Order Routing and Competition in Retail Trade Execution, Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4609895

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI), pp. 1137-1143.

Li, Sida and Ye, Mao and Zheng, Miles, Refusing the Best Price? (December 27, 2020). Available at SSRN: https://ssrn.com/abstract=3763455

Lundberg, S. and S-I, Lee, 2017, A Unified Approach to Interpreting Model Predictions, NIPs Conference Proceedings.

Kwan, Amy, Ronald Masulis, and Thomas McInish, 2015. Trading rules, competition for order flows and market fragmentation. *Journal of Financial Economics* 115, 330-348.

Macey, J. and M. O'Hara, 1997, The Law and Economics of Best Execution, *Journal of Financial Intermediation* 6, 188 -223.

Moinas, S., 2010, Hidden Orders and Liquidity in Order Driven Markets, IDEI Working Paper 600, Institut d'Economie Industrielle (IDEI).

Pan, C., B. Van Ness and R. Van Ness, 2024, Hidden Liquidity Puzzle: Bond ETFs versus Equity ETFs, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4765211.

Philip, R., 2020, Estimating permanent price impact via machine learning, *Journal of Econometrics* 215, 414-449.

Topbas, Yunus and Ye, Mao, 2024, When Exchanges Are Not Exchanges: Evidence From Two Types of Dark Trading Available at SSRN: https://ssrn.com/abstract=4759130

Tuttle, L., 2006, Hidden Orders, Trading Costs and Information, working paper.

Zhu, H., 2014, Do dark pools harm price discovery? *Review of Financial Studies* 27, 747-789.

Appendix A:
Matching Marketable Orders to Limit Orders in the LSEG Trade Data

This Appendix illustrates how it is possible to identify within the LSEG trade data the marketable orders and their corresponding limit orders that give rise to the resulting trade records in LSEG. As noted in the main text, this is made possible by the fact that (a) the LSEG trade data for exchanges is based on individual limit orders that are filled by inbound market orders and (b) LSEG records the time in microseconds reported by an exchange for each executed limit order. As a result, it is possible to reconstruct the market order that triggered the corresponding trades in the LSEG data. For the same reason, it is also possible to reconstruct the top of the order book—displayed and non-displayed—at the time this market order arrived on the venue.

Figures 1A and Figure 2A provide an illustration. In Figure 1A, the blue bars represent the size of the first three price levels for Nasdaq of the displayed offer for shares of Maravai Lifesciences Holdings Inc. (MRVI), a sample security, on May 16, 2023 immediately before 09:50:16.692653 according to the LSEG quote data. At 09:50:16.692653, the LSEG trade data reveals that seven individual buy trades were executed at Nasdaq having this exact timestamp. These trades are shown in Figure 1A as red dots. Note that no corresponding offer depth exists for the two 50 share trades executed at \$15.45 and \$15.47. Additionally, there are two trades at \$15.48—one for 700 shares and one for 50 shares—that exceed the 700 displayed shares on offer at this price. The two trades for 302 shares and 100 shares at \$15.51 likewise exceed the 100 shares displayed at \$15.51. These discrepancies with the order book are explained by the fact that the LSEG trade flags indicate that these three 50-share trades along with the 302-share trade each executed against non-displayed orders. Thus, collectively, these trades reflect a marketable buy order for 452 shares “walking up” the Nasdaq book and executing against displayed and non-displayed orders until the final trade in the sequence is printed at \$15.51 per share for 302 shares.

In Figure 2A, we use this information to reconstruct the offer orders sitting at the top of the Nasdaq book immediately prior to 09:50:16.692653, distinguishing those orders that are displayed from those that are not. Figure 2A highlights the presence of the two aggressively priced 50-share non-displayed orders at \$15.45 and \$15.47, and it also shows how additional non-displayed depth existed at the displayed prices of \$15.48 and \$15.51. The stacked nature of the bar graph also illustrates the execution priority of the two order types insofar that limit orders will be filled from bottom (blue) to top (grey) as a marketable order walks up each price level.

When we apply this technique to all trades in our sample, we find that our 215,417,621 individual trade records can be linked to 151,425,017 marketable orders.

Appendix B:
Estimating the Prevalence of Non-Displayed Orders Beyond the Top of the Order Book

As noted in the main text, it is difficult to estimate the prevalence of NMS Hidden liquidity through an exchange's order book due to the fact that we can observe non-displayed orders in the LSEG data only when a trade happens to interact with a non-displayed order. Moreover, because so few market orders interact with limit orders beyond the top of an exchange's order book, unexecuted non-displayed orders will not appear in our data. We can, however, examine the prevalence of non-displayed orders for those market orders that are sufficiently large that they walk up or down an exchange's order book.

To undertake this analysis, we return to Figures 1A and 2A where we illustrate how we can reconstruct how our sample of 215 million executed limit orders interacted with the market orders that initiated the executed trades. In particular, because LSEG records the time in microseconds reported by an exchange for each executed limit order, it is possible to reconstruct the market order that triggered the corresponding trades in the LSEG data. For the same reason, it is also possible to reconstruct the top of the order book—displayed and non-displayed—at the time this market order arrived on the venue. Using this technique, we find that our 215,417,621 individual trade records can be linked to 151,425,017 marketable orders.

For each of these marketable orders, we additionally assess the displayed price levels at which it interacted. To illustrate, consider again the trades in MVRI shown in Figure 1A. Overall, the marketable order depicted in the figure interacted with four price levels as follows. First, because the two 50-share trades executed at $15.45 and $15.46 were at prices better than the best displayed offer, we classify these trades as interacting with level 0, which indicates a price level superior to the exchange's best displayed bid or offer. Next, the trades priced at $15.48 (the first displayed offer), $15.50 (the second displayed offer), and $15.51 (the third displayed offer) were classified as interacting at price levels 1, 2 and 3, respectively. We conduct an analogous classification for each of the 151 million marketable orders through order level 5 (i.e., the fifth level of displayed prices), classifying any interactions with price levels beyond this level as "level 6." Therefore, for each marketable order, we can identify whether the order exhausted all displayed liquidity through five levels of displayed prices. In the event trades execute against non-displayed orders between two displayed price levels (for example, a hypothetical trade for 10 shares at $15.49 in Figure 1A), we classify such trades at the displayed price level that must be exhausted immediately before any such order can be filled (or level 1 in this hypothetical).

When we examine how the roughly 151 million marketable orders interacted with the 215 limit orders in our sample, we find that the vast majority executed against only one price-level of orders posted at the top of the exchange's order book. Specifically, over 147 million (97%) interacted with just one price level (either level 0 or level 1) in our classification scheme. Thus, just 4,018,520 marketable orders interacted with more than one price level of liquidity, and these orders produced 8.3 million order-level combinations, ranging from 3,782,467 marketable orders that interacted with two price levels (7,564,934 order-level interactions) to 1,613 marketable orders that interacted with 6 or more price levels (11,291 order-level interactions).

Our interest is in understanding how much non-displayed liquidity rests at each order level beyond order level 0; therefore, we focus on the rate at which a marketable order interacted with non-displayed liquidity at each order level beginning with order level 1, conditional on the marketable order exhausting all displayed liquidity at that level. Returning to Figure 1A to illustrate, we can surmise non-displayed liquidity existed at level 1 and level 3 because the marketable order was sufficiently large to exhaust all displayed liquidity at these two levels, and we observe trades executing against non-displayed orders at these two levels. In contrast, Figure 1A reveals that no non-displayed liquidity was available at order level 2 because the order exhausted level 2's displayed liquidity yet no non-displayed orders were executed before the order hit level 3.

We present the results of this analysis in Table 1A. The top half of the table focuses on buy orders, and the first row indicates that when a marketable buy order was sufficiently large to exhaust all displayed offer liquidity at order level 1, a notable 80.4% of these orders hit a non-displayed order at an offer price equal to order level 1 or less than order level 2. By this measure, a marketable buy order that exhausted the liquidity posted at an exchange's best displayed offer would have an 80% chance of hitting a non-displayed offer before interacting with the next best displayed offer posted on the exchange. The subsequent rows, however, indicate that the likelihood of hitting non-displayed offers declines substantially for marketable buy orders that exceeded posted liquidity beyond the exchange's best displayed offer. For instance, when a marketable buy order was sufficiently large to exhaust all displayed offers at order level 2, only 38% of these orders hit a non-displayed offer at a price equal to order level 2 or less than order level 3, while the rate declines to 25% to 28% for marketable buy orders that exhaust liquidity at order levels 3, 4 and 5. Note, however, that the number of instances that a marketable buy order exhausted displayed liquidity beyond an exchange's BBO was quite rare within the sample. Substantially similar results appear for marketable sell orders.

In short, the presence of non-displayed orders declines as a market order walks up or down the order book, suggesting that non-displayed orders are used primarily by traders to place aggressively-priced limit orders.

Figure 1:
Exchange Hidden Order Fills As % of All Order Fills, Jan. 2019 - Mar. 2024

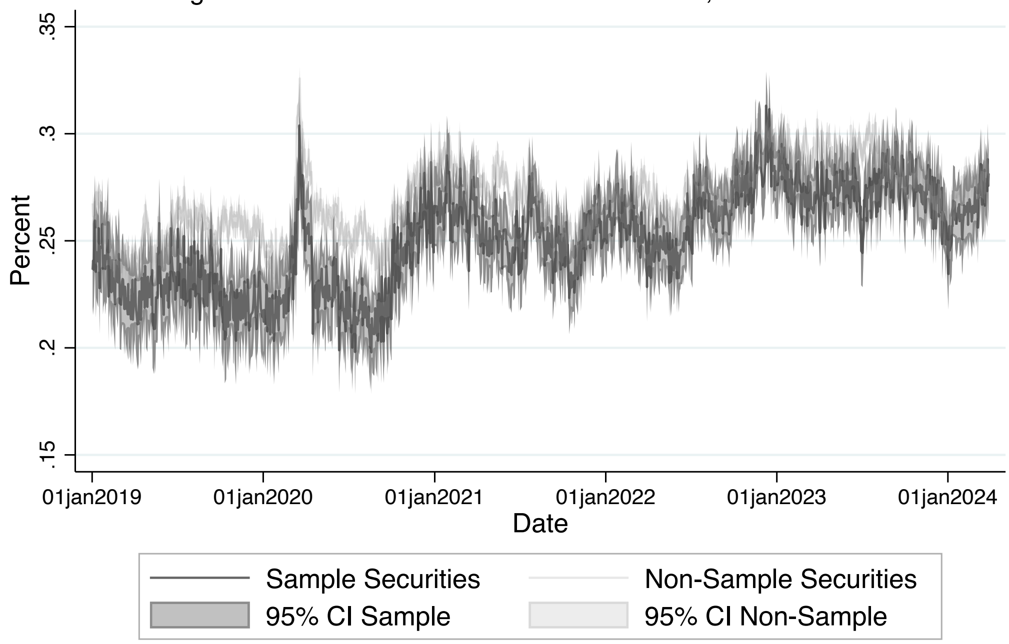
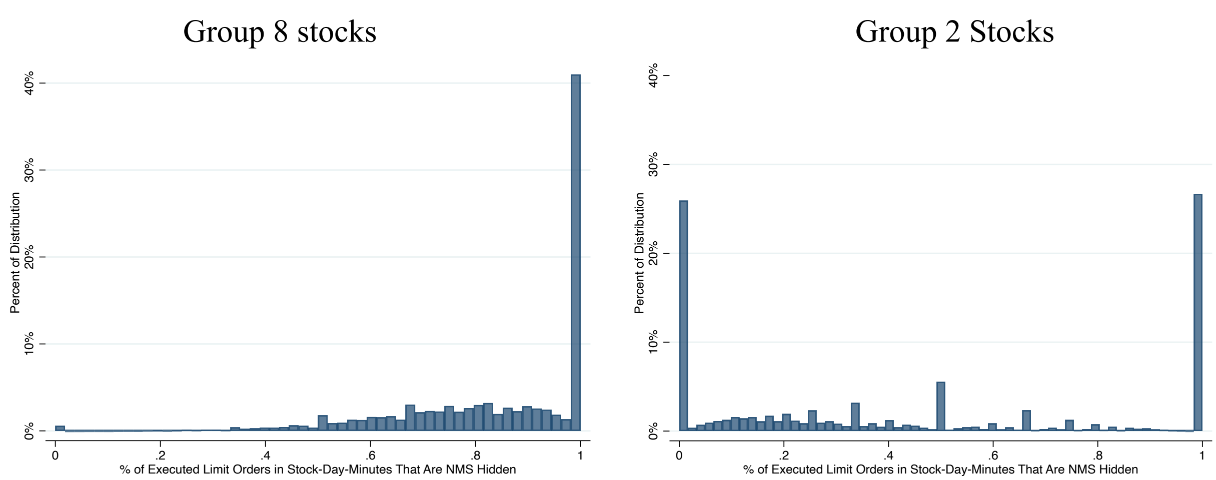


Figure 2: Histogram Showing Percent of Trades Observed in a Stock-Day-Minute Satisfied with NMS Hidden Liquidity



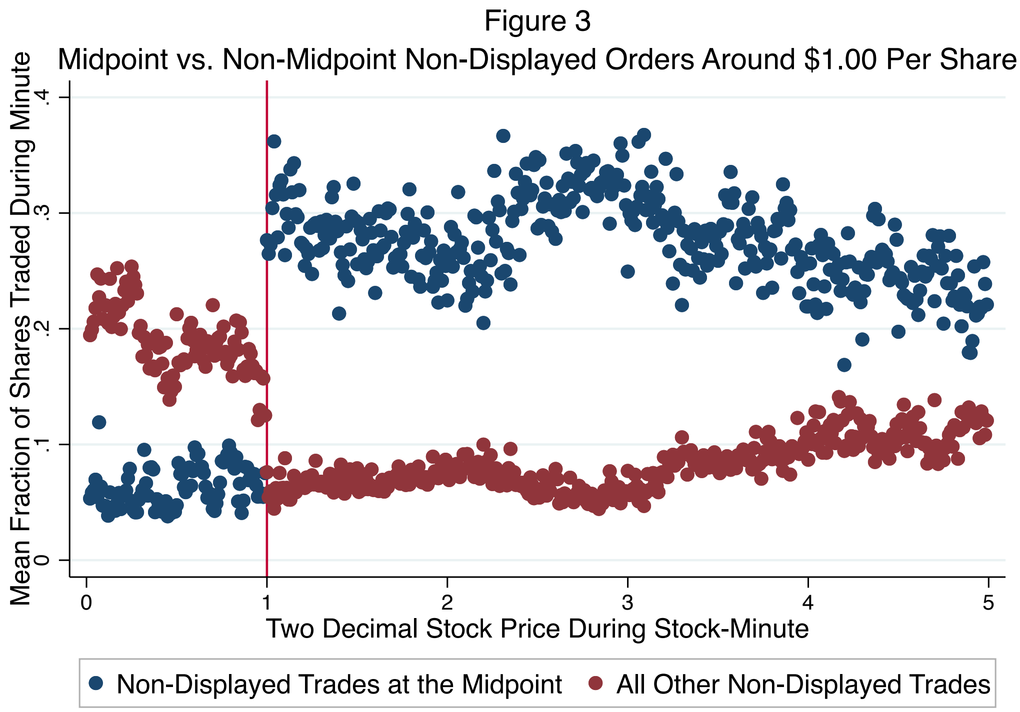


Figure 4: Decomposition of Price Improvement to the NBBO

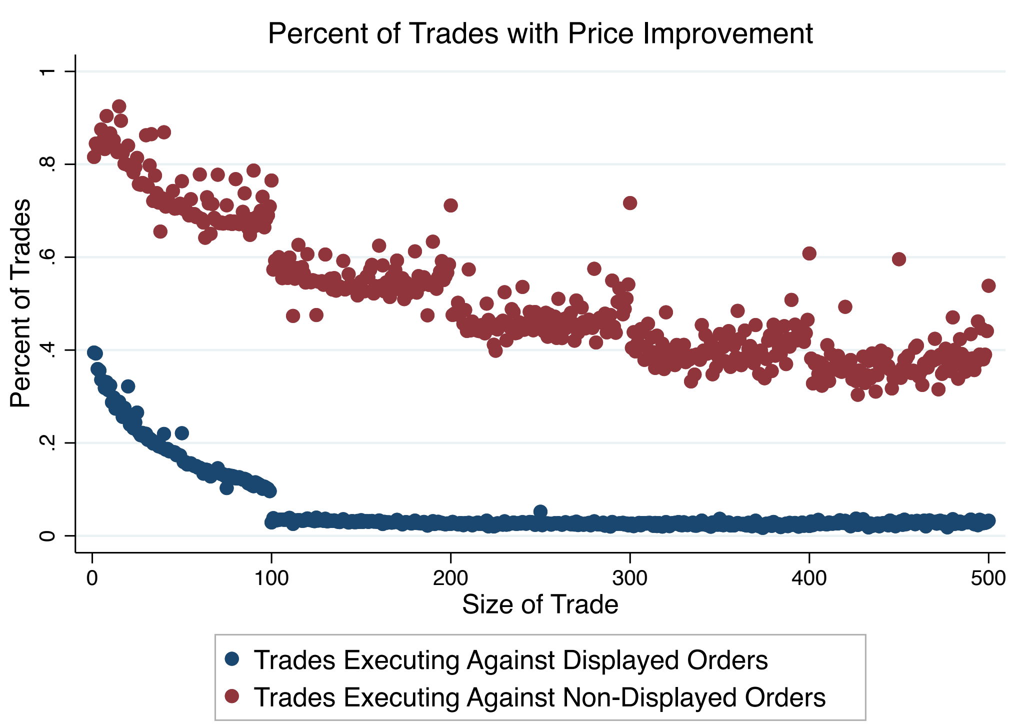


Figure 5: Amount of Price Improvement by Trade Size

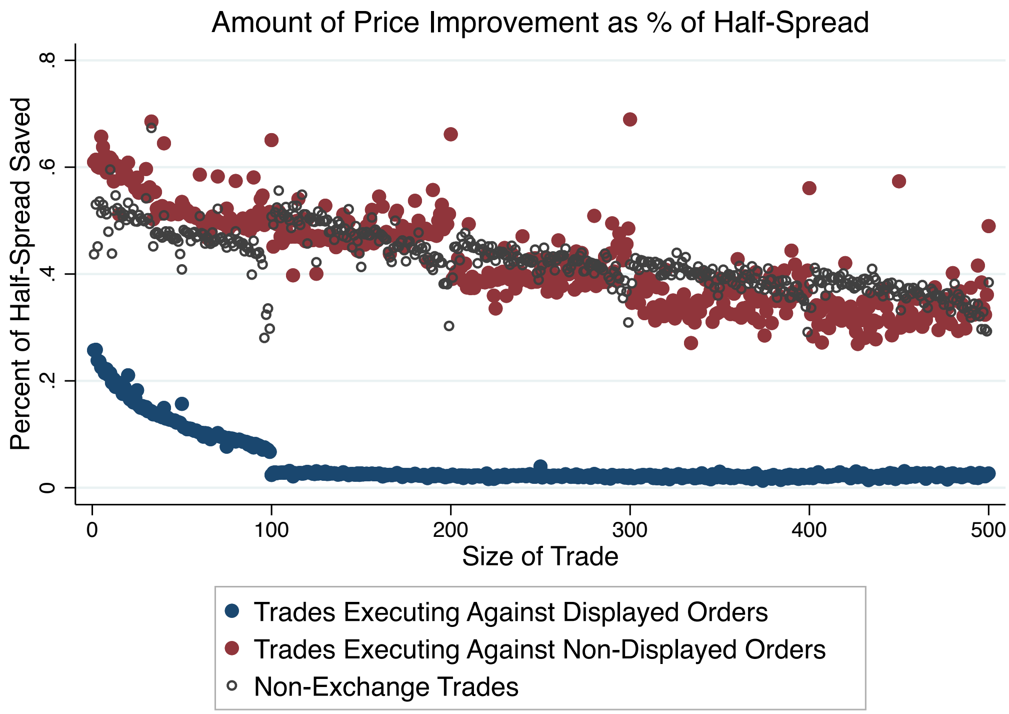


Figure 6: ROC and AUC for XGBoost Implementation of Equation (2)

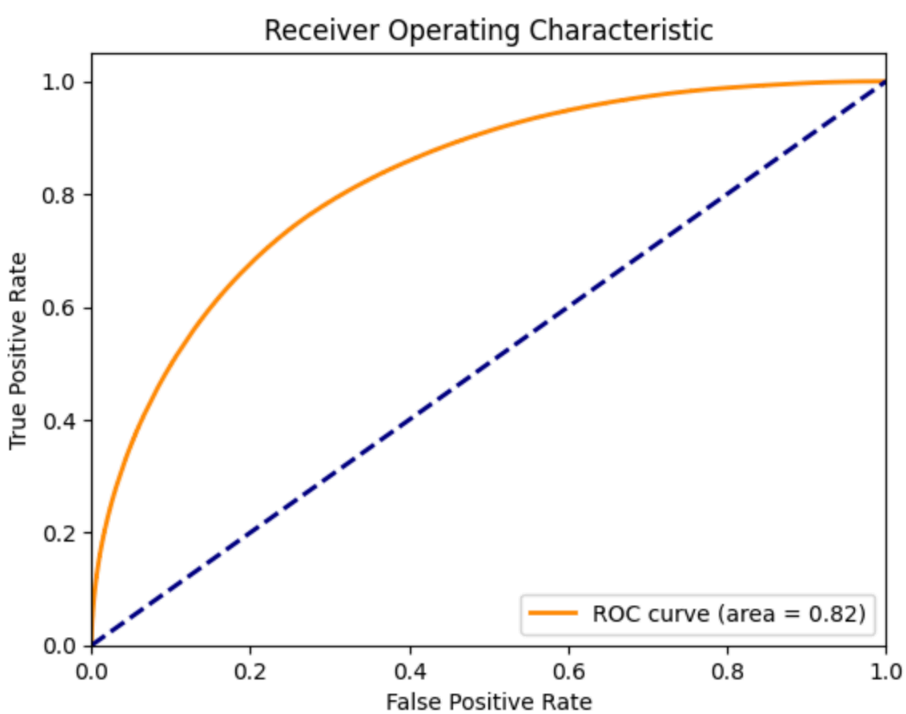


Figure 7: SHAP Values for the Top 20 Features of the Training Data

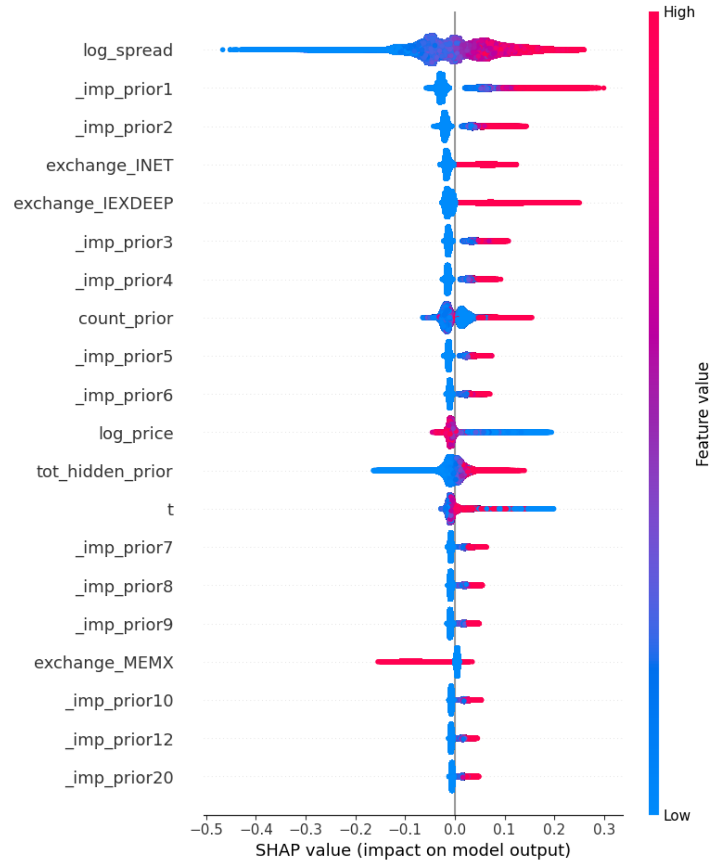


Figure 8: Distribution of Trades Interacting with Non-Displayed Orders By Minute

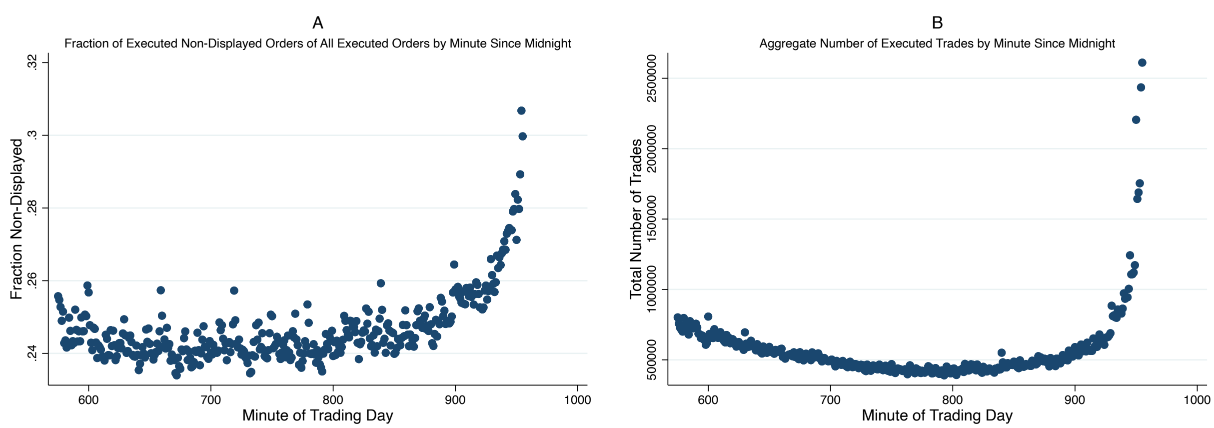


Figure 9: Probability of Hidden Liquidity Across Exchanges for ZWS at 15:50 on June 27, 2023

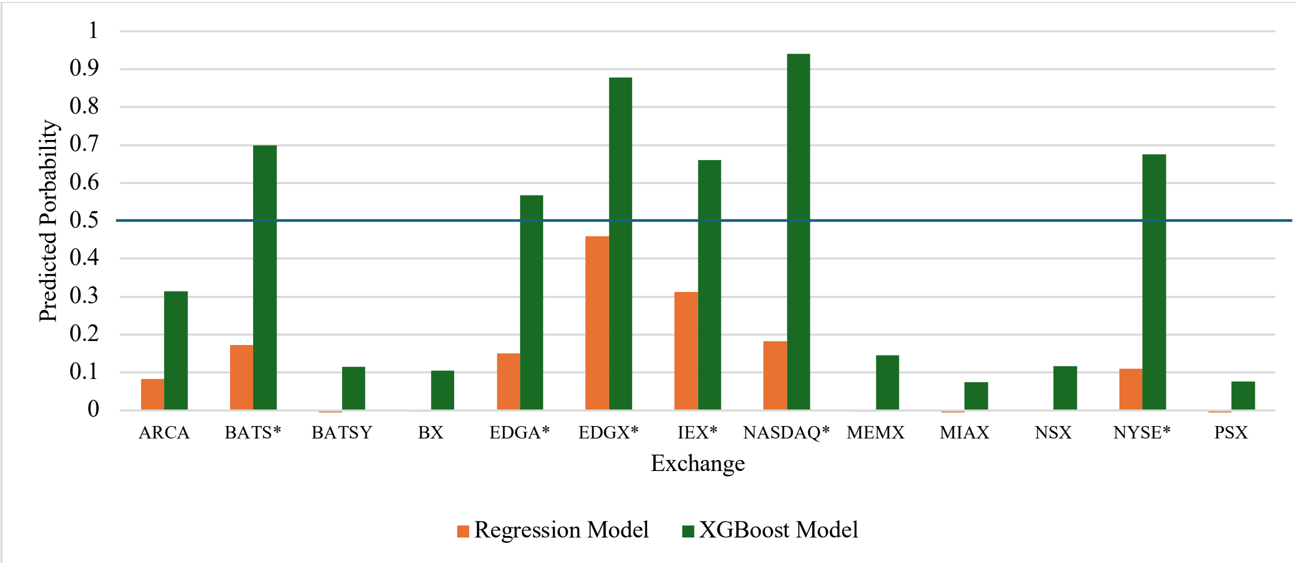


Figure 1A: Ask Depth on Nasdaq for MRVI Compared to Trades on Nasdaq for MRVI

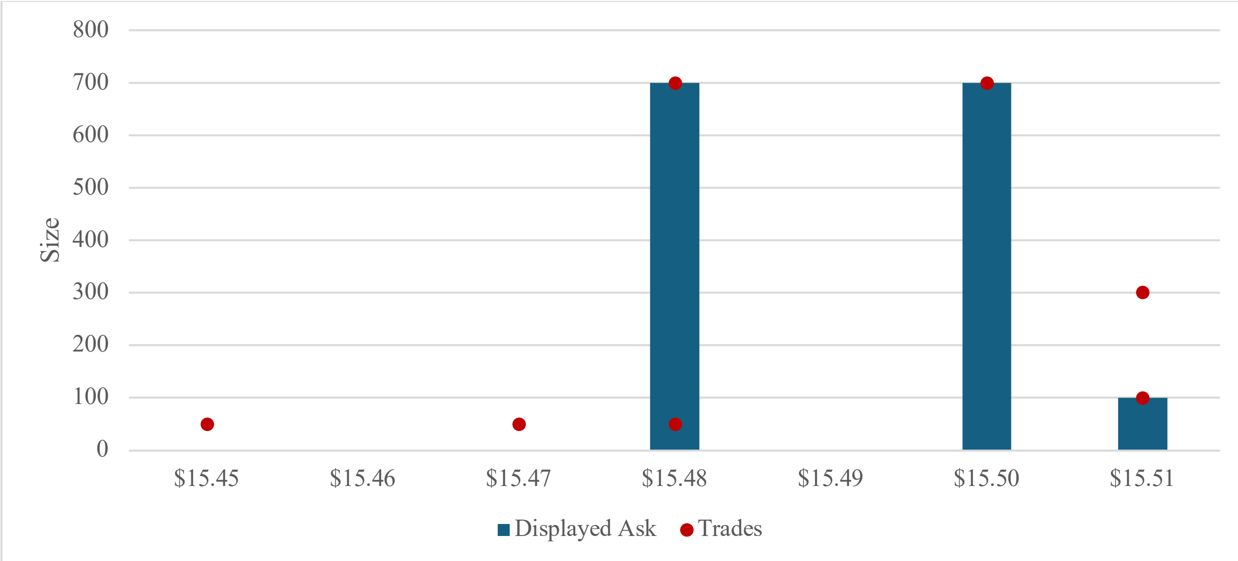


Figure 2A: Ask Depth on Nasdaq for MRVI: Displayed + Non-Displayed Implied by Trades

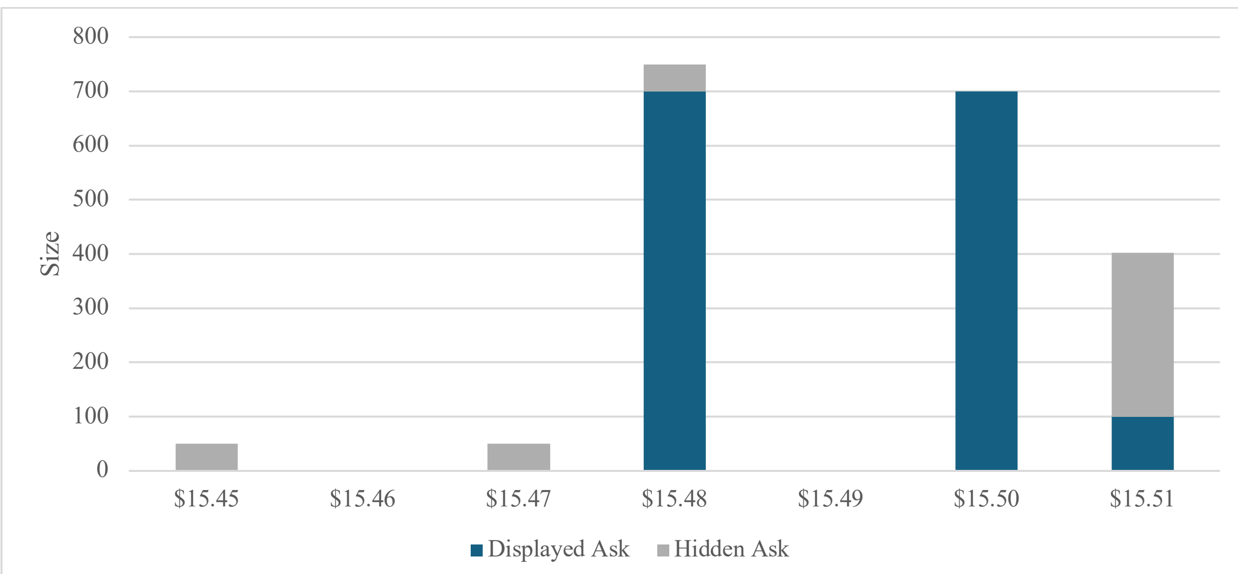


Table 1

This table illustrates how trades in GEF.B appear in the TAQ trade data compared to how they appear in the LSEG trade data. TAQ data appear in columns (1)-(6) and LSEG data appear in columns (7)-(14).

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(9)	(10)	(11)	(12)	(13)	(14)
TAQ Trade Data						LSEG Trade Data						
DATE	TIME	EX	TR SCOND	SIZE	PRICE	Time	source	symbol	shares	price	side	flags
20230103	9:39:02.244710144	N	I	1	79.1	9:39:02.244710	NYSE	GEF.B	1	79.1	B	-
20230103	9:40:58.631925760	N		119	79.1	**9:40:58.631925**	**NYSE**	**GEF.B**	**20**	**79.1**	**B**	**-**
						9:40:58.631925	**NYSE**	**GEF.B**	**99**	**79.1**	**B**	**-**
20230103	9:42:14.005192192	N	F I	1	79.1	9:42:14.005192	NYSE	GEF.B	1	79.1	B	s
20230103	9:45:07.544044544	N	I	10	78.6	**9:45:07.544044**	**NYSE**	**GEF.B**	**9**	**78.6**	**S**	**-**
						9:45:07.544044	**NYSE**	**GEF.B**	**1**	**78.6**	**-**	**h**
20230103	9:47:51.181199360	N		180	78.5	**9:47:51.181199**	**NYSE**	**GEF.B**	**80**	**78.5**	**S**	**-**
						9:47:51.181199	**NYSE**	**GEF.B**	**100**	**78.5**	**S**	**-**
20230103	9:47:51.181199360	N	I	20	78.54	9:47:51.181199	NYSE	GEF.B	20	78.54	-	h
20230103	9:47:51.181216512	N	F I	28	78.5	9:47:51.181216	NYSE	GEF.B	28	78.5	S	s
20230103	9:47:51.181279744	N	I	72	78.5	9:47:51.181279	NYSE	GEF.B	72	78.5	S	-
20230103	9:59:25.002382080	N	I	5	78.31	9:59:25.002382	NYSE	GEF.B	5	78.31	S	-
20230103	9:59:25.002428672	N	I	1	78.31	9:59:25.002428	NYSE	GEF.B	1	78.31	S	-

Table 2

This table provides descriptive statistics of our sample stocks during our sample period of January 3, 2023 through June 30, 2023. *Market Cap* is the average dollar value of each stock's equity (in millions). *Trading Volume* is the average dollar value of trades in each stock (in millions). *Price* is the average closing price for each stock. *# Trades* is the total number of trades in a stock. *Fraction Non-Displayed* and *Fraction Odd* are the fraction of all trades in a stock that were marked as non-displayed by LSEG or that were for fewer than 100 shares, respectively.

Variable	Mean	p25	p50	p75	N
Market Cap	$3,340.0	$87.9	$377.0	$1,990.0	311
Trading Volume	$28.0	$0.7	$4.2	$18.1	311
Price	$38.5	$3.0	$10.8	$37.4	311
# Trades	692,661	77,731	313,304	871,390	311
Fraction Non-Displayed	28.0%	22.1%	27.4%	32.7%	311
Fraction Odd	65.7%	50.5%	71.6%	84.2%	311

Table 3:

This table provides statistics on the aggregate amount of sample trades that executed against NMS Hidden Liquidity versus NMS Displayed liquidity. Panel A analyzes how these two forms of liquidity are reflected in the individual trades reported by LSEG. Panel B analyzes how these two forms of liquidity interacted with the total dollar value of trades in the sample. *Non-Displayed - Round* represents liquidity provided by non-displayed orders with a size of 100 shares or more. *Non-Displayed - Odd* represents liquidity provided by non-displayed orders with a size of less than 100 shares. *Displayed - Round* represents liquidity provided by displayed orders with a size of 100 shares or more. *Displayed - Odd* represents liquidity provided by dispalyed orders with a size of less than 100 shares.

Panel A: Executed Limit Orders

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Non-Displayed		Displayed		Fraction NMS Hidden		
	Total	Round	Odd	Round	Odd	% of Column (1) Non-Displayed	% of Column (1) Displayed Odd	% Both
All	215,414,318	18,673,202	35,138,200	130,263,951	31,338,965	24.98%	14.55%	39.53%
Top of Book	206,894,060	18,520,681	34,853,955	123,422,860	30,096,564	25.80%	14.55%	40.34%
% Top of Book	96.04%	99.18%	99.19%	94.75%	96.04%			

Panel B: Dollar Volume

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Non-Displayed		Displayed		Fraction NMS Hidden		
	Total	Round	Odd	Round	Odd	% of Column (1) Non-Displayed	% of Column (1) Displayed Odd	% Both
All	$467.27	$82.14	$51.37	$290.97	$42.80	28.57%	9.16%	37.73%
Top of Book	$436.54	$80.89	$50.51	$265.19	$39.96	30.10%	9.15%	39.25%
% Top of Book	93.42%	98.48%	98.34%	91.14%	93.35%			

Table 4

This table summarizes the extent to which trades occuring over the course of a stock-day-minute interacted with non-displayed orders and displayed odd lot orders across eight different groups of sample stocks based on the average stock price of trades during the stock-day-minute. Group 1 includes stocks with a price less than or equal to $1.00, and the cutoff points for Groups 2 through 7 are $5.00, $10.00, $25.00, $50.00, $100.00, and $200.00. Stocks with a price higher than $200.00 per share are placed in Group 8. Trades are based on trade records reported by LSEG; therefore, each trade reflects an execution against an individual limit order. Statistics represent the average percent across all stock-days in the sample. Panel A analyzes how the individual trades reported by LSEG interacted with non-displayed orders and displayed odd lot orders. Panel B analyzes the dollar value of trades during a stock-day-minute interacted with these two forms of liquidity.

Panel A: Executed Limit Orders

	(1)	(2)	(3)	(4)	(5)
	% of Limit Orders	Median # of Limit Orders	% Liquidity Provided by Non-Displayed Orders	% Liquidity Provided by Odd Lot Orders	% Both
All Limit Orders					
Group 1: Price<$1.00	100%	6	22.2%	5.9%	28.1%
Group 2: $1<Price<=$5	100%	4	36.4%	7.9%	44.3%
Group 3: $5<Price<=$10	100%	8	32.2%	12.6%	44.7%
Group 4: $10<Price<=$25	100%	14	30.9%	15.9%	46.8%
Group 5: $25<Price<=$50	100%	10	34.2%	25.4%	59.5%
Group 6: $50<Price<=$100	100%	13	38.1%	30.7%	68.8%
Group 7: $100<Price<=$200	100%	22	37.9%	35.1%	73.0%
Group 8: Price>$200	100%	18	45.4%	38.5%	83.9%
Limit Orders at Top of Book					
Group 1: Price<$1.00	98.7%	6	22.6%	6.0%	28.6%
Group 2: $1<Price<=$5	98.6%	4	37.0%	7.9%	44.9%
Group 3: $5<Price<=$10	98.2%	8	32.7%	12.7%	45.4%
Group 4: $10<Price<=$25	98.0%	13	31.6%	16.0%	47.6%
Group 5: $25<Price<=$50	97.6%	9	35.0%	25.5%	60.5%
Group 6: $50<Price<=$100	96.8%	12	39.2%	30.8%	70.0%
Group 7: $100<Price<=$200	96.1%	21	39.1%	35.1%	74.2%
Group 8: Price>$200	95.1%	17	47.1%	38.0%	85.1%

Panel B: Trading Volume

	% of Trading Volume	Median Trading Volume	% Liquidity Provided by Non-Displayed Orders	% Liquidity Provided by Odd Lot Orders	% Both
All Trading Volume					
Group 1: Price<$1.00	100%	$399.2	25.9%	3.7%	29.7%
Group 2: $1<Price<=$5	100%	$615.4	35.8%	5.9%	41.7%
Group 3: $5<Price<=$10	100%	$2,838.1	31.8%	9.2%	41.0%
Group 4: $10<Price<=$25	100%	$9,761.5	31.5%	10.7%	42.2%
Group 5: $25<Price<=$50	100%	$8,784.0	35.2%	18.2%	53.4%
Group 6: $50<Price<=$100	100%	$23,090.2	39.4%	21.0%	60.4%
Group 7: $100<Price<=$200	100%	$75,093.7	39.5%	23.1%	62.6%
Group 8: Price>$200	100%	$116,520.4	46.2%	26.7%	72.9%
Volume at Top of Book					
Group 1: Price<$1.00	98.2%	$389.1	26.5%	3.9%	30.3%
Group 2: $1<Price<=$5	98.2%	$603.5	36.5%	6.0%	42.5%
Group 3: $5<Price<=$10	97.7%	$2,759.7	32.5%	9.3%	41.8%
Group 4: $10<Price<=$25	97.4%	$9,441.6	32.4%	10.9%	43.2%
Group 5: $25<Price<=$50	96.8%	$8,359.7	36.3%	18.5%	54.8%
Group 6: $50<Price<=$100	95.9%	$21,589.5	40.8%	21.2%	62.0%
Group 7: $100<Price<=$200	95.1%	$69,698.1	41.0%	23.3%	64.3%
Group 8: Price>$200	94.0%	$107,184.6	48.3%	26.5%	74.9%

Table 5:

This table summarizes the extent to which sample trades during a stock-day-minute executed against either non-displayed limit orders or odd lot limit orders priced better than the NBBO. *Percent Non-Displayed* represents the fraction of all trades that executed against non-displayed limit orders priced better than the prevailing NBBO. *Percent OL* represents the fraction of all trades that executed against displayed odd lot limit orders priced better than the prevailing NBBO. *Percent Both* is the combined total of Percent Non-Displayed and Percent OL. *Percent Midpoint* represents the fraction of all trades that executed against non-displayed orders priced at the midpoint of the NBBO. In all cases, the NBBO used is the NBBO that prevailed at the time of the trade as reported by LSEG. Trades are based on trade records reported by LSEG; therefore, each trade reflects an execution against an individual limit order. Stocks are grouped based on the daily stock price using the same groups described in Table 4. Statistics represent the average percent across all stock-days in the sample. Panel A examines the rate at which individual trades executed against these limit orders. Panel B examines the dollar volume of trades that executed against these orders.

Panel A: Executed Limited Orders

	(1)	(2)	(3)	(4)
	Percent Non-Displayed	Percent OL	Percent Both	Percent Midpoint
Group 1: Price<$1.00	13.5%	5.9%	19.5%	6.4%
Group 2: $1<Price<=$5	32.9%	7.9%	40.8%	28.9%
Group 3: $5<Price<=$10	29.1%	12.6%	41.6%	22.7%
Group 4: $10<Price<=$25	27.5%	15.9%	43.4%	20.6%
Group 5: $25<Price<=$50	32.0%	25.4%	57.3%	19.1%
Group 6: $50<Price<=$100	35.9%	30.7%	66.6%	16.5%
Group 7: $100<Price<=$200	35.9%	35.1%	71.0%	13.3%
Group 8: Price>$200	44.3%	38.5%	82.8%	12.1%
All	31.1%	20.0%	51.1%	19.5%

Panel B: Dollar Volume

	(1)	(2)	(3)	(4)
	Percent Non-Displayed	Percent OL	Percent Both	Percent Midpoint
Group 1: Price<$1.00	13.6%	3.7%	17.3%	6.3%
Group 2: $1<Price<=$5	32.0%	5.9%	37.9%	28.0%
Group 3: $5<Price<=$10	28.6%	9.2%	37.8%	22.3%
Group 4: $10<Price<=$25	28.0%	10.7%	38.7%	21.2%
Group 5: $25<Price<=$50	32.8%	18.2%	51.0%	20.3%
Group 6: $50<Price<=$100	36.9%	21.0%	57.9%	18.1%
Group 7: $100<Price<=$200	37.0%	23.1%	60.2%	15.2%
Group 8: Price>$200	44.9%	26.7%	71.6%	13.8%
All	31.5%	13.8%	45.3%	20.1%

Table 6

The following table provides by exchange the average percent of trades across all stock-exchange-minutes that executed against non-displayed orders priced better than the NBBO at the beginning of the minute. The table also includes the mean number of total trades across all stock-exchange-minutes.

	(1)	(2)	(3)	(4)	(5)	(6)
	All		Price>$100		Price>$1 & Spread<.02	
Exchange	% Non-Displayed	# of trades	% Non-Displayed	# of trades	% Non-Displayed	# of trades
Maker-Taker Exchanges						
AMEX	15.26%	2.0	6.15%	1.7	21.47%	1.9
ARCA	17.01%	6.4	28.24%	5.8	10.57%	6.9
BATS	22.47%	5.6	31.94%	5.9	17.98%	5.9
CHX	9.59%	1.5	12.00%	1.6	2.79%	1.6
EDGX	22.05%	5.2	27.85%	4.2	20.24%	5.5
IEX	49.18%	3.6	41.04%	4.5	52.91%	3.8
NASDAQ	30.66%	11.8	36.35%	16.0	29.42%	13.1
MEMX	7.33%	4.7	16.23%	3.1	2.72%	5.4
MIAX	5.34%	3.0	19.16%	2.0	1.04%	3.1
NYSE	19.16%	7.0	30.49%	9.5	11.43%	7.0
PSX	13.26%	2.2	24.81%	2.1	4.62%	2.1
Taker-Maker Exchanges						
BATSY	22.76%	2.7	46.86%	2.3	8.25%	2.5
BX	25.29%	2.3	41.59%	2.4	9.59%	2.2
EDGA	23.81%	2.9	54.21%	2.8	7.68%	3.0
NSX	15.96%	2.6	36.75%	2.1	6.06%	2.7
All	23.6%	5.8	32.5%	6.5	17.4%	6.0

Table 7

This table presents regression estimates for models predicting whether price-improving non-displayed quotes (PINQs) are present during a stock-exchange-minute. All estimates are based on data assigned to the "training" period defined as the first four months of the sample period. Coefficients for stock exchange fixed effects are displayed first, followed by *Price* (the natural log of the value-weighted average price from the prior trading minute), *Spread* (the natural log of the quoted spread at the end of the prior minute), *Penny* (an indicator for whether the quoted spread is less than 2 cents) as well as the interaction of *Penny* with exchange, *Return* (the return on stock *i* over the prior minute), *Count* (the number of trades in stock *i* on exchange *x* during the prior minute), *Total_Count* (the aggregate number of trades in stock *i* on exchange x during the prior minute), *Total_ND* (the aggregate number of trades interacting with non-displayed orders in stock *i* on exchange x during the prior minute). All models include stock, date and minute-of-the-day fixed effects and, where indicated, twenty lagged values of the dependent variable. Robust standard errors (in brackets) are double-clustered by stock and day.

	(1)	(2)
DV:	Percent PIHQ	Percent PIHQ
ARCA	0.0423**	-0.0345**
	[0.0194]	[0.0158]
BATS	0.0848***	-0.00229
	[0.0191]	[0.0153]
BATSY	0.209***	0.156***
	[0.0238]	[0.0214]
BX	0.176***	0.124***
	[0.0212]	[0.0189]
CHX	-0.108***	-0.106***
	[0.0165]	[0.0151]
EDGA	0.230***	0.142***
	[0.0267]	[0.0229]
EDGX	0.0570***	-0.00925
	[0.0178]	[0.0151]
IEXDEEP	0.333***	0.114***
	[0.00998]	[0.0114]
NASDAQ	0.144***	-0.0106
	[0.0179]	[0.0150]
MEMX	-0.0397**	-0.0649***
	[0.0176]	[0.0155]
MIAX	-0.0393**	-0.0434**
	[0.0186]	[0.0171]
NSX	0.136***	0.107***
	[0.0210]	[0.0192]
NYSE	0.0845***	-0.0086
	[0.0191]	[0.0158]
PSX	0.0624***	0.0488**
	[0.0204]	[0.0187]
Price	-0.0691***	-0.0673***
	[0.00723]	[0.00616]
Spread	0.0965***	0.0867***
	[0.00574]	[0.00517]
Return	0.575***	0.646***
	[0.0747]	[0.0644]
Count	-0.000559***	-0.000578***
	[9.59e-05]	[0.000101]
Total_Count	-0.000192***	-5.66e-05***
	[2.60e-05]	[1.06e-05]
Total_ND	0.00122***	0.000320***
	[0.000118]	[4.21e-05]
Penny*AMEX	0.0885***	0.0540***
	[0.0207]	[0.0170]
Penny*ARCA	-0.0693***	-0.0276
	[0.0204]	[0.0166]
Penny*BATS	-0.0476**	-0.0258
	[0.0223]	[0.0167]
Penny*BATSY	-0.309***	-0.250***
	[0.0246]	[0.0225]
Penny*BX	-0.274***	-0.208***
	[0.0221]	[0.0198]
Penny*CHX	-0.00395	0.0117
	[0.0222]	[0.0207]
Penny*EDGA	-0.329***	-0.239***
	[0.0283]	[0.0243]
Penny*EDGX	-0.0142	-0.0107
	[0.0234]	[0.0178]
Penny*IEXDEEP	-0.0211	-0.00377
	[0.0195]	[0.0137]
Penny*NASDAQ	-0.0183	-0.000893
	[0.0252]	[0.0168]
Penny*MEMX	-0.0728***	-0.0414**
	[0.0196]	[0.0175]
Penny*MIAX	-0.0969***	-0.0732***
	[0.0211]	[0.0195]
Penny*NSX	-0.249***	-0.205***
	[0.0227]	[0.0209]
Penny*NYSE	-0.121***	-0.0546***
	[0.0221]	[0.0174]
Penny*PSX	-0.174***	-0.140***
	[0.0212]	[0.0195]
Constant	0.685***	0.647***
	[0.0405]	[0.0355]
Observations	21,227,215	21,227,215
R-squared	0.180	0.231
Firm FE	Yes	Yes
Lagged DV	No	Yes
Date FE	Yes	Yes
Minute FE	Yes	Yes

Table 8

This table presents perforance statistics for three models predicting whether price-improving non-displayed quotes (PINQs) are present during a stock-exchange-minute in the test data, defined as stock-exchange-minute observations occuring in the last two months of the sample period. Columns (1) & (2) present performance statsitics for the regression models described in Table 7. Column (3) presents performance statistics for XGBoost models using the same parameters set forth in Columns (2) of Table 7.

	(1)	(2)	(3)
	Equation 1 (no lags): Regression	Equation 1 (with lags): Regression	Equation 1 (with lags): XGBoost
Accuracy	0.5993	0.6358	0.7488
Precision	0.6198	0.8137	0.7183
Recall	0.0763	0.1530	0.6445
F1	0.1359	0.2576	0.6794

Table 1A

This table presents estimates of the probability that a market order will find non-displayed liquidity at each level of displayed prices on an exchange, conditional on the order exceeding all displayed liquidity at that price level.

Marketable Buy Orders:

Price Level	Mean	SD	N
1	80.4%	39.7%	2,603,492
2	39.3%	48.8%	113,474
3	28.2%	45.0%	30,128
4	24.5%	43.0%	11,998
5	25.1%	43.4%	5,409

Marketable Sell Orders:

Price Level	Mean	SD	N
1	78.1%	41.4%	2,588,907
2	38.3%	48.6%	124,971
3	28.5%	45.1%	33,222
4	26.3%	44.0%	13,151
5	27.0%	44.4%	6,089