NBER WORKING PAPER SERIES

SEQUENTIAL SEARCH FOR CORPORATE BONDS

Mahyar Kargar
Benjamin Lester
Sébastien Plante
Pierre-Olivier Weill

Working Paper 31904
http://www.nber.org/papers/w31904

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2023

We thank Briana Chang (discussant), Tatyana Deryugina, Terry Hendershott, Tim Johnson, Jian Li (discussant), Dmitry Livdan, Yi Li (discussant), Maureen O'Hara, Christian Opp (discussant), Gábor Pintér (discussant), Kerry Y. Siani (discussant), Alan Shepard, Kumar Venkataraman (discussant); conference participants at AEA/CEANA, 18th Central Bank Conference on the Microstructure of Financial Markets, FIRS, NBER Asset Pricing Meeting, NBER Big Data and Securities Markets, SED Annual Meeting, Stern/Salomon Microstructure Conference, Tsinghua SEM Conference on Finance and Development, West Coast Search and Matching Workshop; and seminar participants at Bank of Canada, Drexel, Federal Reserve Bank of Minneapolis, Federal Reserve Bank of Philadelphia, Federal Reserve Bank of Richmond, INSEAD, Paris Dauphine University, Princeton University, UC Berkeley Haas, UC Irvine, UCL, UCLA, UC Riverside, UIUC, USC, Wharton, and University of Wisconsin-Madison for comments and suggestions. Lian Chen provided expert research assistance. The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. All errors are our own responsibility. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Sequential Search for Corporate Bonds
Mahyar Kargar, Benjamin Lester, Sébastien Plante, and Pierre-Olivier Weill
NBER Working Paper No. 31904
November 2023
JEL No. D83,G0,G10,G12

ABSTRACT

In over-the-counter (OTC) markets, customers search for counterparties. Little is known about this process, however, because existing data is comprised of transaction records, which are only informative about the end of a successful search. Leveraging data from the leading trading platform for corporate bonds, we offer evidence about the search process: we analyze customers' repeated attempts to trade (successful and unsuccessful). We estimate that it takes two to three days to complete a transaction after an unsuccessful attempt, with substantial variation depending on trade and customer characteristics. Our analysis offers insights into the sources of trading delays in OTC markets.

Mahyar Kargar
Gies College of Business
Department of Finance
University of Illinois at Urbana-Champaign
435 Wohlers Hall
1206 S. Sixth Street
Champaign, IL 61820
kargar@illinois.edu

Benjamin Lester
Research Department
Federal Reserve Bank of Philadelphia
10 Independence Mall
Philadelphia, PA 19106
Benjamin.Lester@phil.frb.org

Sébastien Plante
Wisconsin School of Business
University of Wisconsin
5182 Grainger Hall
Madison, WI 53706
splante@wisc.edu

Pierre-Olivier Weill
Department of Economics
University of California, Los Angeles
Bunche Hall 8283
Los Angeles, CA 90095
and NBER
poweill@econ.ucla.edu

1 Introduction

Over-the-counter (OTC) markets play a key role in the U.S. financial system. They include most fixed income securities, asset-backed securities, repurchase agreements, and various types of derivatives, along with a significant fraction of equity trading volume (Weill, 2020). Unlike exchange-based markets, OTC markets are decentralized: participants must find a willing counterparty and agree on the terms of trade. The prevailing wisdom is that there are frictions in this process and, as a result, completing a trade often takes time.

However, in contrast to other frictional markets, such as the labor or housing markets, little is known about the process that unfolds while an investor searches for a suitable offer. The reason is simple: existing data from OTC markets is comprised of transaction records, which contain information about the time and price at which a trade occurred, but not about the time that investors spent searching. Hence, empirical estimates of the time it takes to trade in OTC markets—along with an understanding of the sources of these trading delays—have remained elusive.

In this paper, we leverage a proprietary data set that allows us to unpack the sequential search process of investors in one of the most studied OTC financial markets—the market for U.S. corporate bonds. The data provides a complete record of all inquiries made by customers and the corresponding replies from dealers on the leading electronic trading platform for corporate bonds, MarketAxess (MKTX). We supplement this data with transactions recorded by the Trade Reporting and Compliance Engine (TRACE), which allows us to find inquiries that failed to result in trade on the electronic platform but were ultimately successful through other channels (i.e., "voice" trades). Crucially, by observing both successful *and unsuccessful* inquiries, the data allows us to estimate how long it takes a customer to trade and how this length of time depends on the characteristics of both the order and the customer. Moreover, by studying the behavior of customers and dealers over the course of the search process, our analysis offers new insights into the sources of trading delays.

At first glance, it might seem counterintuitive to use data from an electronic platform to study the magnitude and sources of trading delays in OTC markets. After all, one of the primary reasons for introducing an electronic platform is to *eliminate trading frictions*. We argue that this data set is a natural starting point for several important reasons. First, at a practical level, it is the only data source with direct observations of the time that investors spend trying to execute a particular trade, the various strategies they employ over the search process, and the corresponding behavior of dealers in response to repeated inquiries. Moreover, investors typically use MKTX for the most liquid trades—namely, smaller quantities of investment grade bonds—so that our estimates can be seen as a natural lower bound on time-to-trade. Finally, since electronic platforms eliminate many obvious physical barriers to finding a counterparty (such as the time-consuming process of sequentially contacting dealers), the discovery of trading delays within a market with an electronic platform offers a window into the deeper frictions underlying the sequential search process.

Our analysis proceeds in three steps. First, we document that executing a trade often requires multiple attempts or "inquiries," and thus organize the data in a way that captures the sequential nature of investors' search process. For example, we find that inquiries fail to result in trade about 30% of the time, which is consistent with the findings of Hendershott and Madhavan (2015) from an earlier period. We go beyond this earlier work by following customers after a failed inquiry: by combining data from MKTX and TRACE, we can observe when customers make additional electronic inquiries on MKTX for the same trade (successfully or not); when they complete the trade outside of the electronic platform (via voice channels); and when they alter their trading strategy (by attempting to trade a different quantity or a different bond) or abandon the trade altogether.

Second, exploiting the granularity of the data through the lens of a structural model, we use maximum likelihood to derive novel estimates of time-to-trade in the corporate bond market. Our estimates reveal the observable characteristics that affect time-to-trade, including the trade direction (buy or sell), the trade size, the type of bond, and the characteristics of the customer making the inquiry. For a fairly liquid trade—namely, an odd-lot purchase of an investment-grade bond by a customer who is "well connected" to dealers—we find that it takes approximately two days to complete a purchase after an initial inquiry fails. Block trades (with size above $5 million) take about one day longer than micro-size trades (with size below $100,000), while bonds with amounts outstanding below the median take half a day more to trade. We find that trade is about twice as fast when customers want to sell, relative to when they want to buy. Moreover, we find significant heterogeneity across customers as they attempt to trade. In particular, customers vary widely in the number and quality of replies they receive in response to an inquiry, which we call their "connectedness," and this has a significant impact on their time to trade: it takes nearly twice as long for customers in the bottom 70% of our connectedness measure to trade, relative to those customers in the top decile. Comparing time to trade on MKTX against traditional voice-based methods, we observe that orders trade faster on MKTX for all size categories except block trades, suggesting customers use the platform for execution quality rather than price discovery.

The probability that an initial inquiry fails also varies systematically across observable attributes of the request: for example, the failure rate is approximately 14% for inquiries made by the most connected customers to complete the fairly liquid purchase request described above, and rises to more than 50% for the least connected customers attempting to execute more difficult trades. Hence, the expected time to trade from the moment an initial inquiry is placed ranges from as little as a few hours to several days. Of course, this represents a lower bound on the total time-to-trade, since we do not have information about how long the customer was searching prior to submitting their first inquiry on MKTX.[1]

Our estimates of time-to-trade are helpful for at least two reasons. First, they can be directly applied to quantitative analyses based on search-theoretic models, since the arrival rates we estimate are crucial, yet controversial inputs that are typically identified via indirect inference. Second, the correlations we find between our estimates of time to trade and other observables provide a natural starting point for additional empirical and theoretical work. For example, our findings suggest that heterogeneity in customers' ability to elicit offers from dealers is crucial for understanding OTC market outcomes. As a second example, our finding that it takes longer for customers to buy a bond than it does to sell it suggests that it might be important to understand frictions in the process of dealers actually finding (or "sourcing") a bond before selling it to a customer. Alternatively, this finding could be indicative that customer-sellers are more distressed than customer-buyers, on average. Distinguishing between these two explanations, and incorporating the subsequent findings into existing theoretical frameworks, is crucial for understanding, e.g., the effects of various shocks or policy proposals on prices and allocations in OTC markets.

In the third step of our analysis, we dig deeper into the behavior of customers and dealers over the course of the search process in order to shed light on the underlying economic mechanisms that generate trading delays in OTC markets. After all, the existence of an electronic trading platform enables customers to reach a wide set of dealers with a click of the button. If one were take a narrow, literal view of trading frictions, they might anticipate that an electronic trading platform would make the US corporate bond market operate much like a frictionless exchange. Yet we find that it does not: inquiries regularly fail to generate trade; both the number and quality of offers vary significantly across customers and, for a given customer, over the course of their search; and, ultimately, a trade often takes a nontrivial amount of time to execute. These observations raise

[1]For example, the customer could have searched for this trade on voice before the initial MKTX inquiry, or could have been consulting dealer "runs" to determine when it would be worthwhile to make an inquiry. See Hendershott, Li, Livdan, Schürhoff, and Venkataraman (2021) for a more detailed description of the dissemination of dealer runs.

a number of key questions. Why are customers rejecting dealers' offers and sending an identical inquiry hours (or days) later? Why do some dealers reply to one inquiry and not the other? Why do some dealers change their offer from one inquiry to the next?

We find that, not surprisingly, an important motivation for repeated inquiries is to find a better quote. As in the classic model of McCall (1970), both the number and the quality of dealers' replies vary over time, so that customers who reject an offer and continue to search typically trade at a better price. Among those customers who eventually trade, we estimate that an additional inquiry results in an average spread improvement of approximately 3 bps, which is about 18% of the average spread in our sample. Next, we find that a key reason for variation in the number and quality of quotes appears to derive from *fluctuations in dealers' inventory holdings*. In particular, we document that dealers are more likely to reply to a customer's request to purchase a bond—and more likely to offer a better ask price—when the dealer sector as a whole is holding a relatively large quantity of the bond in inventory.[2]

We also document a relationship between the number of inquiries a customer has made and the number and quality of dealers' replies—a form of duration dependence that could potentially be explained by dealers learning over time (e.g., the so-called "ringing phone curse"). However, we find that this relationship more likely reflects unobserved heterogeneity across the orders being requested. Specifically, we analyze the number and quality of dealers' replies by adding fixed effects for each sequence of repeated inquiries by the same customer for the same bond. This allows us to control for both observed and unobserved heterogeneity during the course of the search process. We find that outcomes become much more stable, suggesting that the effects of dealers learning over the course of a search are either minimal or offset by other forces in this setting.

Related literature

Our work is most closely related to the few other papers that have used the proprietary data from MKTX to analyze the impact of electronic trading on corporate bond market conditions (e.g., Hendershott and Madhavan, 2015; O'Hara and Zhou, 2021; Hendershott, Livdan, and Schürhoff, 2021). Our analysis differs from these papers in both our focus and our approach. To start, rather than treating each inquiry as an independent observation, we organize the data into clusters of inquiries that are made within a short time period by the same customer to buy/sell the same bond;

[2]Our findings are symmetric for customer requests to sell: fewer dealers reply, and the quality of their bid offers are worse, when inventory holdings are relatively high.

this allows us to document a number of new insights about the sequential nature of customers' search process. Second, with the aid of a structural model, we use maximum likelihood to derive novel estimates of time-to-trade conditional on the characteristics of the trade and the customer. Lastly, we exploit a number of unique features of the data to explore the underlying sources of trading delays, which have yet to be explored in the existing literature.

Our work also contributes to the vast empirical literature that studies corporate bond market liquidity based on transaction data. Some prominent examples include Schultz (2001), Bessembinder, Maxwell, and Venkataraman (2006), Edwards, Harris, and Piwowar (2007), Goldstein, Hotchkiss, and Sirri (2007), Bao, Pan, and Wang (2011), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), and many others.[3] Our main contribution relative to this literature is to document what happens *before* a trade occurs, i.e., to unpack the search process of customers in the US corporate bond market and shed light on the frictions that prevent immediate trade. To the best of our knowledge, our paper is among the first to derive *direct* empirical estimates of time to trade—a key dimension of liquidity in the corporate bond market, and a crucial input into search-theoretic models of OTC markets. Hendershott, Li, Livdan, and Schürhoff (2020) pursue similar goals but for a different dimension of liquidity (the cost of trade failures) in a different market (the market for collateralized loan obligations).

Our attempt to measure time-to-trade is related to earlier work in the OTC literature. In particular, since existing data provides detailed information about dealers' transactions, several authors have used the restrictions of search models to identify dealers' time-to-trade. For example, according to the models of Afonso and Lagos (2015), Üslü (2019), and Brancaccio and Kang (2021)—in which search is random and the distribution over agents' state is continuous—every meeting results in a trade, allowing them to identify time-to-trade from the observed frequency of dealers' transactions. While this identification strategy may be reasonable for dealers, it is problematic for customers who presumably spend long periods of time out of the market: clearly, observing that a customer trades once a year does not imply that it takes a year to find a counterparty. Hugonnier, Lester, and Weill (2020) also use a structural model, along with data on asset turnover and the length of intermediation chains, to identify dealers' time-to-trade (with other dealers and with customers separately); however, this approach does not allow for the identification of *customers'* time-to-trade. Gavazza (2016) is able to measure customers' time-to-trade in a structural model by taking advantage of aggregate information about the total number of real assets (in his

[3] See Bessembinder, Spatt, and Venkataraman (2020) for a survey.

case, aircraft) for sale at a time. Such information is typically not available in OTC market data. More recently, Pintér and Üslü (2021) offer an indirect measurement of customers' time-to-trade based on joint observations of trade size and frequency.[4] Our contribution is to propose a more direct approach, based on granular observations, which does not rely on the restrictions imposed by a specific structural model.

Finally, our approach is related to the large literature that attempts to estimate the key objects of interest in the standard sequential search model of McCall (1970), which was first used in financial economics by Garbade and Silber (1976). Early attempts to do so in a labor market context include Kiefer and Neumann (1979) and Flinn and Heckman (1982), among others. As in labor economics, this simple partial equilibrium model is a natural starting point for interpreting micro data, as it helps rationalize failed inquiries, repeated attempts to trade, and price dispersion. However, while we find it useful to formulate a search-theoretic model to motivate our empirical exercise and interpret its findings, it is important to note that our measurement does not impose theoretical restrictions from the model. In this way, our findings are related to the seminal work of Lo, MacKinlay, and Zhang (2002), which also relies on survival analysis to measure time-to-trade in a central limit order book.[5] Our analysis differs along several dimensions due to different trading mechanisms and data structures. In addition, the magnitudes are vastly different: their findings from equities three decades ago are in the order of minutes, whereas our findings from the current corporate bond market are measured in hours and days.

The remainder of the paper has three parts. Section 2 describes the data. Section 3 provides theory and evidence about time to trade. Finally, Section 4 empirically analyzes some likely sources of trading delays.

2 Data

Our main source of data is MarketAxess (MKTX), the leading electronic trading platform in the corporate bond market. Prior to the introduction of MKTX in 2000, the corporate bond market operated almost exclusively under a "voice-based" trading system, whereby customers would sequentially contact dealers (via telephone or chat) one at a time to solicit a quote. Stepping into

[4] Pagnotta and Philippon (2018) offer a more detailed discussion of trading speeds across various markets and trading mechanisms.

[5] Deville and Riva (2007) also apply survival analysis in a different context to measure the time it takes for arbitrage opportunities to close in option markets.

this market, MKTX offered a trading platform allowing buy-side traders (henceforth customers) to query multiple dealers at once via an electronic request for quote (RFQ), thus reducing the time-consuming process of gathering quotes and potentially increasing competition across dealers.

As of the third quarter of 2022, MKTX accounted for approximately 21% of total trading volume in the corporate bond market.[6] However, as Table 1 reports, the market share of MKTX, calculated as the ratio of trading volume executed on MKTX to the total trading volume observed in TRACE, varies significantly across trade size and bond rating categories. In particular, consistent with earlier findings by O'Hara and Zhou (2021), customers tend to use MKTX more intensely for smaller trades and for investment-grade (IG) bonds. Indeed, as one can see in column (2), more than half (54%) of the volume for odd lot trades (with size between $100,000 and $1 million) of IG bonds is executed through MKTX.

When requesting a quote on the MKTX platform, customers specify the bond they wish to trade, the desired quantity, the trade direction or "side" (buy or sell), and the duration of the auction (usually between 5 and 20 minutes). Once submitted, an inquiry is sent to a customer's list of pre-authorized dealers.[7] On the receiving end, dealers observe the details of the inquiry, including the customer's identity. The receiving dealers may respond to the inquiry with a quote, but are not obligated to do so. At the end of the auction, customers observe the terms of the replies (if any), and can choose to either accept one (and only one) of the offers or reject them all.[8]

Our sample from MKTX covers all trading activity from January 3, 2017 to March 31, 2021. The data contain detailed information on customer inquiries, dealer responses, and customer trading decisions. More specifically, for each inquiry, we observe the submission time (stamped at the second), an anonymized customer identifier, the CUSIP (Committee on Uniform Securities Identification Procedures) number of the bond, the quantity requested, the trade side (buy or sell), the number of dealers who received the request, and several other attributes. For every response to an inquiry, we observe the anonymized identifier of the responding dealer together with their quote. For inquiries that result in a transaction, we observe the time at which trade occurs and the terms

[6]Source: MarketAxess quarterly report for 2022Q4, available from: https://investor.marketaxess.com.

[7]Starting in 2012, MKTX initiated Open Trading, an all-to-all trading option that allows other investors, including other customers and non-pre-authorized dealers, to respond to customer RFQs. However, as Hendershott, Livdan, and Schürhoff (2021) report, the vast majority of trades are still intermediated by a dealer.

[8]The main variation in dealers' offers is price. In principle, dealers can respond to an offer with a different quantity, but in practice more than 97% of dealer responses are at the quantity level requested by the customer.

of trade. Note that we observe all inquiries, including those that do *not* result in a trade, either because the inquiry receives no responses or because the customer rejects all responses.

Importantly, when an inquiry fails to trade on MKTX, a customer may trade outside the platform, either via bilateral trades with dealers, or via other electronic platforms such as Tradeweb or Bloomberg. In what follows, we will for simplicity refer to the trades occurring outside of MKTX as "voice trade."[9] As we describe in greater detail below, we attempt to identify these trades using the enhanced version of the TRACE data set provided by FINRA, which contains detailed reports of every successful trade, whether it has an electronic or voice origin. When working with TRACE, we filter the data following the standard procedure laid out in Dick-Nielsen (2014), and merge the cleaned data with the Mergent Fixed Income Securities Database (FISD) to obtain bond fundamental (e.g., credit rating, amount outstanding, coupon rate, and so on). Following the bulk of the academic literature, we exclude variable-coupon, convertible, exchangeable, and puttable bonds, as well as asset-backed securities, privately placed instruments, and foreign securities, both in the TRACE and MKTX data. We also exclude primary market transactions.

Finally, we measure trade execution costs as a markdown or markup relative to a mid-point price that we calculate using benchmark bid and ask prices provided by MKTX. In particular, MKTX uses a proprietary algorithmic pricing engine for corporate bonds called Composite+ (or "CP+"), which outputs reference bid and ask prices at a high frequency (every 15 to 60 seconds).[10] These forecasts can be used to benchmark a significant fraction of TRACE records: 95% (90%) of TRACE records for investment-grade (high-yield) bonds can be matched to a standing CP+ forecast.

[9]The trades we refer to as "voice" are much more likely to be bilateral trades with dealers than electronic trades. For a back-of-the-envelope calculation, we note that, according to Coalition Greenwich, by the end of 2022, electronic trades had market share about 40% in IG (and 33% in HY). Recall that MKTX market share is 20% in IG (16% in HY). This implies that 75% (= $(100-40)/(100-20)$) of IG and 80% (= $(100-33)/(100-16)$) of HY trades that we refer to as "voice trade" are bilateral trades with dealers.

[10]The construction of the forecasts follows two steps. First, MKTX trains a machine learning (ML) algorithm using three distinct sources of bond trading data: (1) historical TRACE prints; (2) indicative bond price data streamed by dealers; and (3) request for quote responses sent by liquidity providers on the MKTX trading platform. Beyond trading data, MKTX uses bond level information and other broad market data, such as CDX levels, to train the prediction engine. The engine is recalibrated overnight at a daily frequency. Second, the calibrated engine is used over the next trading day to generate real-time reference bid and ask prices of individual bonds using all available intraday information. For more details about CP+, see https://www.marketaxess.com/price/composite-plus.

2.1 The query process: parent and child orders

To give the reader a sense of how the query process works—and to motivate the way we organize and analyze the data—we first provide a few (representative) examples of inquiries. To start, panel (a) of Table 2 provides an example of a successful inquiry. In this example, a customer submitted an inquiry to buy \$300,000 in par value of an investment-grade bond issued by Bank of America. The customer received six replies from dealers, whose anonymized identifiers are provided in column (6). Note that, because the bond in question is investment-grade, dealer responses in column (7) are expressed in terms of yield spread relative to a benchmark Treasury bond, so that a higher yield spread implies a lower purchasing price. As we can see from this column, dealers' quoted yield spreads vary between 126.37 and 129.70 basis points. In the second row of column (9), the entry "Done" shows that the customer accepted the best (highest) offer.[11]

Panel (b) of Table 2 provides an example of an unsuccessful inquiry. This inquiry was submitted by the same customer and for the same bond as the inquiry reported in panel (a), but two days later. This time, the customer requested to purchase \$490,000 in par value instead of \$300,000. Nine dealers responded to the customer's new request. By comparing the identifiers of responding dealers for both inquiries, we see that five of the six dealers who responded to the first inquiry also responded to the second inquiry. Four additional dealers, who had not replied to the first inquiry, replied to the second inquiry. However, the customer decided to pass on the best offer (a yield spread of 127.01 bps), as indicated by the "did not trade" (DNT) flag in the last column.

While customer inquiries are informative about the trading process in and of themselves, a careful examination of the data reveals that individual inquiries are often part of larger trading orders. As a result, individual inquiries should not always be treated as independent observations. To help the reader see why, Table 3 reports *all* the inquiries that the customer in our previous examples submitted to purchase this particular Bank of America bond over a six month period. To save space, we do not report the responses that each inquiry received, and report only whether or not a given inquiry resulted in a trade (see column 7). Note that the first and second inquiries reported in Table 3 correspond to the inquiries reported in panel (a) and panel (b) of Table 2.

Notice immediately that the customer made repeated *successful* purchase inquiries for the same bond over an eight day period. Of the six inquiries, four were successful and led to the purchase of

[11] In the last row of column (9) in Table 2, the entry "Cover" identifies the second best offer. MKTX informs dealers who submit the second best offer of the rank of their quote. Dealers who submit lower-ranked offers do not learn their relative position in the auction.

300, 490, 290, and 680 bonds (with \$1,000 par value) for a total of 1,760 bonds. This anecdotal evidence suggests that customers sometimes execute large orders by submitting a sequence of smaller inquiries, i.e., they split their trades. The second noteworthy feature of Table 3 is that the customer twice followed an *unsuccessful* inquiry by resubmitting an identical inquiry (same bond, quantity, and trade side) soon afterward; this phenomenon is observed after both the second and fourth inquiries. While both of these unsuccessful inquiries received multiple dealer responses, the customer chose to pass. Hence, the example in Table 3 suggests that even when customers are able to simultaneously contact a large number of dealers, they may choose to turn down the offers they receive and query the market again at a later time.

As we discuss in detail below, splitting a larger order into smaller inquiries is a regular features of the data. Hence, we argue that a natural first step is to organize inquiries into clusters, representing the total quantity of a particular bond that a customer is attempting to trade, which we call "parent" orders. Within each parent order, we further partition the set of inquiries into sets of "child" orders in which the customer requests a specific quantity of the bond.[12] In the example above, as one can see in columns (8) and (9) of Table 3, all six inquiries make up a single parent order—where the customer attempts to trade 1,760 units of this particular bond over an eight day period—and this parent order is split into four smaller child orders.

Since the data itself does not explicitly identify parent and child orders, we use the following classification procedure. First, to construct parent orders, we group all inquiries made by a specific customer for a given bond and trade side until we do not observe a new inquiry with the same characteristics (customer, bond, trade side) for N_p days since the last inquiry. The time cutoff N_p is admittedly arbitrary; we set $N_p = 5$ days in our main specification but we obtain qualitatively similar results with other cutoffs (e.g., $N_p = 10$).

Second, we construct child orders by looking at repeated inquiries from a given customer for the same bond, the same trade side, *and the same requested quantity*. We consider all inquiries with these characteristics as part of the same child order until either (i) the most recent inquiry of the child order led to an electronic trade on MKTX; (ii) the customer submitted a new inquiry requesting a different quantity, in which case we initiate a new child order with the updated quantity; or (iii) there is no new inquiry with the same characteristics (same customer, bond, trade side, and trade size) for more than N_c days, where $N_c \leq N_p$. When no new inquiry has been submitted for

[12] We borrow the parent and child order terminology from the equity market literature on institutional trading where large (parent) orders are often split into smaller (child) orders for execution.

more than N_c days, we consider the execution of the child order unsuccessful on MKTX. Here again, the threshold N_c is arbitrary. While we use a cutoff of five days in our main specification, our main results are not sensitive to this choice.

There are two reasons why a child order may be unsuccessful on MKTX. First, the customer may alter their inquiry (by changing the requested trade size) or give up on the trade entirely. Second, the customer might trade the bond via voice. These two outcomes have different economic implications and should be distinguished. Ideally, we would match customer inquiries on MKTX that result in a voice trade using the corresponding TRACE record. However, since TRACE does not report customer identities, it is impossible to match a child order that is traded via voice to its corresponding TRACE record with certainty. Fortunately, this issue can partially be overcome since most corporate bonds trade only a few times a day or less. As a result, the likelihood that two different customers would trade the same quantity of the same bond in the same direction within a few days is arguably low. We thus infer the occurrence of a voice trade by verifying if there exists a record in TRACE with the same characteristics as the unsuccessful child order (same bond, traded quantity, trade side) within five days of that child order's last inquiry on MKTX. In the rare cases where there are multiple matches, we select the closest one in time.

2.2 Summary statistics

Parent and child orders with multiple attempts to trade, such as those described in Table 3, are fairly common in our sample. Panel (a) of Table 4 reports that multiple trading attempts are observed in 26% of parent orders and account for 43% of trading volume. A parent order is considered to have multiple attempts to trade if it is composed of a child order with multiple trading attempts, or if it is composed of multiple child orders. At the child order level, panel (b) of Table 4 reports that multiple trading attempts are observed in 13% of child orders and account for 14% of trading volume. A child order is considered to have multiple attempts to trade if it is composed of two or more inquiries on MKTX, or if it has a single failed inquiry on MKTX followed by a voice trade.[13]

We could, in principle, conduct our analysis at the level of either parent or child orders. However, the splitting of a parent order could be driven by a variety of frictions beside search, including

[13]Our definition of child order is quite restrictive since we require that inquiries have the exact same requested quantity to be considered part of the same child order. Loosening this definition to allow some tolerance (e.g., that a follow-up inquiry is within 10% of the quantity requested in the initial inquiry) increases the number of multi-attempt inquiries. As a result, the numbers reported in Table 4 should be understood as conservative estimates of the prevalence of child orders with multiple attempts to trade.

the mitigation of "information leakage" (as in, e.g., Kyle, 1985). For this reason, we study the sequential search process using child orders as our main unit of observation. Intuitively, focusing on child orders is tantamount to studying the search process for a marginal unit of asset. Another advantage of conducting our analysis at the child order level is that we can use trade quantity to identify those child orders that are eventually traded by voice. If we were to use parent orders instead, we would have to rely on a less stringent criterion, which would introduce measurement errors. Given this choice, our estimates of the time required to trade each child order are a natural lower bound on the time required to trade the full amount a customer wishes to transact.

The child order sequence of events. By construction, the first event that we observe in any child order is an inquiry on MKTX, which either results in a trade or fails. If the initial inquiry fails, and for every failed inquiry thereafter, the next element of the child order is one of four possible events. First, the customer may make another inquiry on MKTX that fails to produce a trade. Second, the customer may make another inquiry on MKTX that results in a trade with one of the dealers that responded. Third, we may find that the customer traded the desired bond-quantity pair outside of the MKTX platform, via voice trade, within a short period of time. Fourth, the customer may give up on this specific trade and "exit," either by sending an inquiry for a different amount or by abandoning the trade altogether. Figure 1 illustrates a child order event tree. Note that we can measure the time elapsed between any two events in this tree, unless it is an exit.

Summary statistics at the child order level. Our focus on child order sets us apart from previous studies, such as Hendershott and Madhavan (2015) or O'Hara and Zhou (2021), who consider the universe of all inquiries and/or trades on MKTX. A simple way to illustrate the conceptual difference between child orders and inquiries is to calculate trade probabilities. Column (1) in Table 5 reports the probability that an inquiry successfully results in trade, whereas column (2) reports the probability that a child order successfully results in trade. Recall that a child order can be successful either at the first inquiry or later in the search process (on MKTX or via voice); columns (3) and (4) report the probability of these two outcomes. The first row of Table 5 reports these statistics in the full sample, and the ensuing rows report the corresponding probabilities for different sub-samples of the data to illustrate the determinants of successful RFQs.

In the full sample, we find that approximately 70% of all inquiries result in a successful trade, which is consistent with the findings of Hendershott and Madhavan (2015) from an earlier time

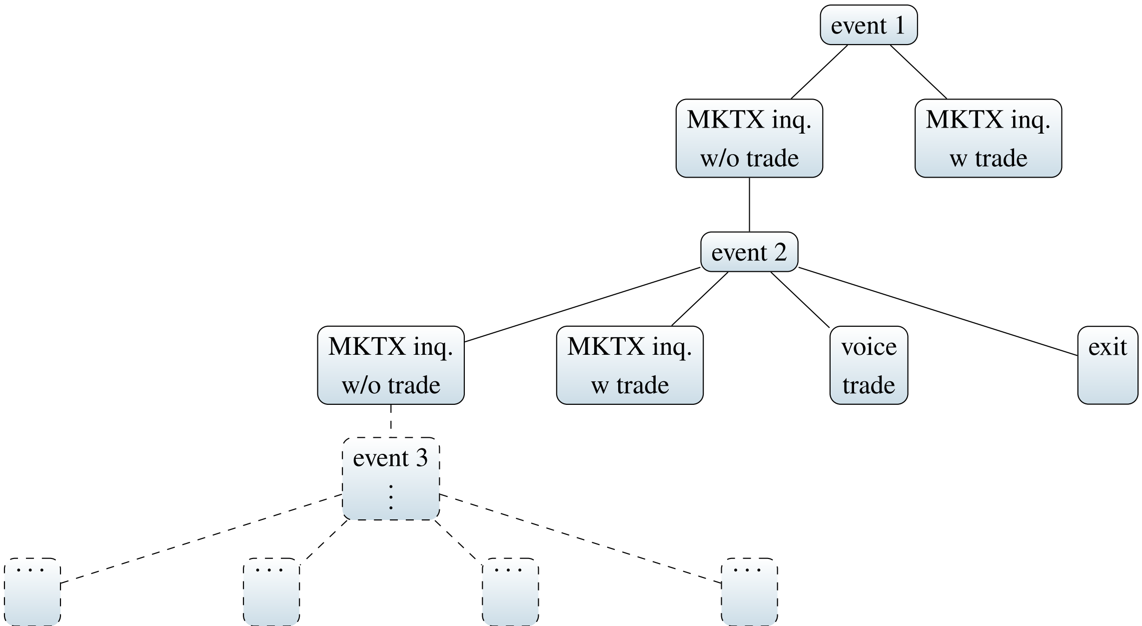


Figure 1. A child order event tree
A child order can be viewed as a sequence of events. Each element of the sequence is one of four possible events: a MKTX inquiry without trade, a MKTX inquiry with trade, a voice trade and, if the child order ends without a trade, an exit. By construction, the first event is always an inquiry on MKTX, either without or with trade.

period. As we discuss in more detail later in the text, approximately 10% of the failed inquiries received no responses and the remaining 90% receive one or more replies. Note that, since child orders can include repeated inquiries on MKTX or via voice, they are naturally associated with larger trading probabilities than inquiries alone. The difference is economically significant: in the full sample, approximately 84% of child orders are eventually fulfilled.

The probability of trade also differs systematically with the properties of the trade and the customer. For example, customer requests to sell a bond are more likely to succeed than requests to buy. Requests to trade bonds with higher ratings, more turnover, more amount outstanding, and less time-to-maturity are also more likely to succeed. Micro size trades (less than $100,000) are also more likely to be fulfilled relative to larger trades.

However, perhaps the most significant source of heterogeneity across trade requests derives from differences across customers. To illustrate this fact, we create a measure of the requesting customer's "connectedness" to proxy for the number of existing relationships with dealers (or other unobserved characteristics that influence the number and quality of replies a customer receives).

To do so, we first regress the average number of dealer responses elicited by a particular customer, controlling for the customer's average inquiry size, the fraction of requests that were sell vs. buy, and the fraction of requests that were for investment-grade vs. high-yield bonds. We then rank customers into deciles based on residuals of this regression. One can see that the most connected customers are much more likely to trade—both at the inquiry and the child order levels—relative to their less connected counterparts.

In order to isolate the effects of specific trade or customer characteristics on trading probabilities, we perform two logistic regressions. In the first regression, the dependent variable is whether trade occurred on MKTX at the inquiry level, while in the second regression the dependent variable is whether trade occurred on MKTX or voice at the child order level in column (2). The independent variables are indicator functions for the customer and trade characteristics described above, along with an indicator to distinguish between the "Covid period" of March 2020, when the corporate bond market suffered a severe disruption, and "normal times" (outside of the Covid period).

We define our "baseline" as a fairly liquid request: an odd lot, investment-grade, buy request from a customer in the top decile of our connectedness measure, requested during normal times (i.e., not Covid), when the bond being requested had above-median turnover and amount outstanding and below-median time to maturity. The results are in Table IA.1 and summarized in Figure 2, which shows the implied trading probabilities from the regression as we vary different attributes of our baseline request. The blue bars represent inquiry-level trade probabilities (on MKTX) and the combined blue and gray bars represent child-order trade probabilities (on MKTX or voice).[14] The figure highlights, again, the important distinction between inquiries and child orders. It also reinforces the message that the "least connected" customers, defined as those in the bottom seven deciles of our connectedness measure, trade with significantly lower probability than those who are better connected.[15] However, these customers frequently continue to search after a failed inquiry, as reflected by the difference between the probability of success at the inquiry level (about 48%,) and the child order level (about 74%).

Finally, Figure 2 also highlights that the probability of trade fell at both inquiry and child order

[14]Keeping in mind that the unit of the coefficients is the log odds ratio of trade, the intercept in column (1) shows that the odds ratio of trade for the baseline category is 2.231, so that the probability that an inquiry in our baseline category is successfully completed is $e^{2.231}/(1+e^{2.231}) = 90\%$ as shown in the second blue bar in the Figures.

[15]The empirical relevance of connections in OTC market has been studied before. For example, Afonso, Kovner, and Schoar (2014) show that connections impact terms of trade in the Federal Funds Market during stressful times. In the UK government bond market, Kondor and Pintér (2022) proxy for the arrival of private information using time variation in the number of connections, measured as the number of dealers a given client trade in a given day.

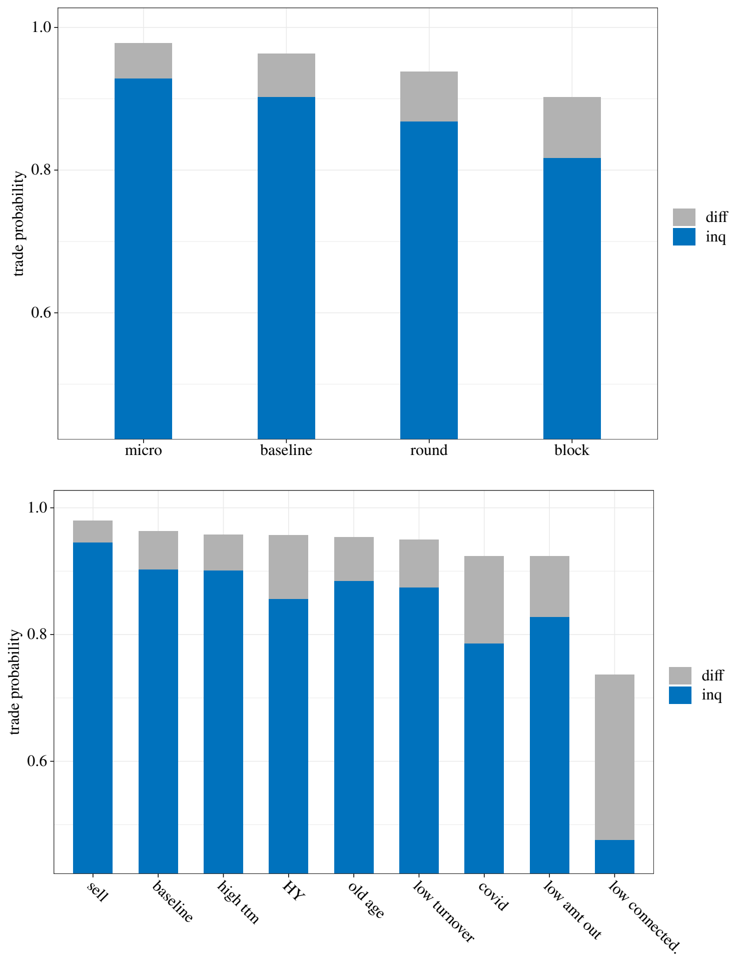


Figure 2. Estimated trade probability on MKTX at inquiry and child order levels
This figure compares the estimated trade probability using logit regression estimates from Table IA.1. The blue bars present trade probabilities (on MKTX) at the inquiry level. The gray bars shows the extra trade probability (on MKTX or voice) for a child order, taking into account the option to make repeated inquiries and trade on voice. The top panel shows trade probabilities for different size categories. The bottom panel presents trade probabilities for non-size categories. Indicators for size and non-size categories are defined in Tables 5 and 8. The baseline category is an odd-lot purchase of an investment-grade bond, with high turnover, low time-to-maturity, and high amount outstanding, during normal times, for a connected investor.

levels during the COVID-19 crisis in March 2020, but that the fall was much less dramatic at the child order level: at the inquiry level, the trade probability falls to about 79%, but at the child order level it falls much less, to 92%. This suggests that sequential search becomes more prevalent during stressful events: it is harder for customers to obtain good quotes on MKTX, but they can compensate by waiting.

One may wonder whether our sample significantly differs from its inquiry- or trade-level counterparts in other dimensions as well.[16] Table IA.2 in the Internet Appendix shows that this is not the case: child-order and inquiry-level summary statistics are broadly the same for trade direction and size and bond characteristics. As in previous studies, we find that trade sizes on MKTX are smaller and bond credit risk is lower than in the market at large. To measure the inter-arrival time of trading opportunities, we will need to restrict the sample further to child orders with at least one failed inquiry. Column (3) of Table IA.2 shows that the summary statistics remain similar, though the sample is now more selected towards high-yield bonds and inquiries of larger size since both are less likely to trade at the first MKTX inquiry.

Table 6 offers additional information about what happens after each failed inquiry on MKTX. For example, the second row shows that if the first inquiry fails, the probability that the following event is a failed inquiry on MKTX is 0.16, the probability that there is a successful inquiry on MKTX is 0.09, the probability that there is a voice trade is 0.26, and the probability that the customer adjusts or abandons the order is 0.48. We will argue later that these probabilities are not straightforward to interpret because of competing risk and selection biases. Notwithstanding these issues, there are a few takeaways from Table 6.

First, a customer's search for a counterparty is often sequential: the probability of failing an inquiry is nontrivial and, conditional on failure, customers often submit repeat inquiries. Second, a customer's search is also nonexclusive: if the first inquiry fails, the child order may eventually trade on MKTX or voice.[17] The third takeaway is that, after a failed inquiry, a customer is fairly likely to end a child order, either by altering the parameters of the inquiry or abandoning the trade altogether. Fourth, the summary statistics in Table 6 show that the frequency distribution over the four events depends on the number of failed inquiries—a form of duration dependence.

Finally, table IA.3 in the Internet Appendix presents inter-arrival times between events in child orders. For example, after one failed inquiry, the average time to the next traded inquiry on MKTX is 0.65 business days (the clock we use to measure business days only runs when the market is open, accounting for the fact that the market is closed at night and during weekends and holidays).

[16]For example, suppose that high-yield bonds trade after twice as many inquiries as investment-grade bonds. Then we would find that the number of high-yield inquiries is twice that of high-yield child orders.

[17]While the probability of a voice trade is larger than that of an MKTX trade, the ratio is not as large as the relative volume of voice to MKTX volume. This suggests that, although trade is nonexclusive, the customers in our sample are using MKTX more intensely than the general population.

However, as we argue below, this estimate is clearly biased downwards, since observing this event requires that none of the other events occur first.

3 Measuring Time to Trade: Theory and Evidence

In this section, we first formulate and solve a sequential search model of a child order in the style of McCall (1970), which was first applied to financial markets by Garbade and Silber (1976). This theoretical detour serves two purposes. First, it clarifies several important issues that arise when interpreting the data; in particular, our derivations below elucidate two sources of bias in estimating time to trade—competing risk and selection—and hence motivate the statistical model we estimate later. Second, it illustrates how our empirical estimates shed new light on search-based models of OTC markets: our analysis offers guidance for the quantitative values of key parameters and highlights which dimensions of the model fit the data well, and which

Second, we formulate a statistical framework inspired by the the structure of the sequential search model and use maximum likelihood to estimate the time it takes for customers to contact and trade with dealers (via MKTX and via voice). This allows us to lay out, for the first time, a set of stylized facts regarding time-to-trade based on direct observations. Note that, at this stage, we remain agnostic about *why* it takes time to trade; we return to this question later, in Section 4, where we offer some empirical evidence about the sources of trading delays.

3.1 A McCall (1970) model of a child order

Time is indexed by $t \in [0, \infty)$. We consider a child order to sell one unit of a perpetual par bond, i.e., a perpetuity with a coupon rate equal to the interest rate, r (the analysis of a purchase is symmetric). We assume that the seller is risk-neutral with discount rate r and values the bond below its par value of 1. Specifically, when she holds the bond, she derives a flow utility $r - c$ for some distress cost $c > 0$. The seller recovers from distress with intensity γ. Upon recovering, we assume that the seller's continuation value reverts to the par value of the bond, she stops searching, and exits the market. In the data, the seller may exit for a variety of other reasons. For instance, this could happen if she updates the quantity requested or makes an inquiry for a different bond. This can be captured by assuming that the continuation value of exiting is different from the par value of the bond. As shown in Appendix A.2, while some details of the analysis change, the main results are upheld.

Consistent with the child order tree of Figure 1, we take $t = 0$ to represent the time at which the seller makes her first inquiry on the electronic market. If the first inquiry is unsuccessful, the seller makes inquiries on the electronic or the voice market with Poisson intensities λ_e and λ_v, respectively. After an inquiry in the electronic market, the seller receives $j \in \{0, 1, 2, \ldots\}$ offers with probability q_j. We represent an offer as a bid $1 - m$, where m is the markdown over the bond par value of 1. We assume further that each offered markdown is drawn independently according to the cumulative distribution function (CDF) $G_e(m)$. Correspondingly, when she makes an inquiry in the voice market, the seller receives just one offer, drawn according to the CDF $G_v(m)$. For simplicity we assume that, for both distributions, the lower bound of the support is 0. As will be clear below, the optimal trading strategy of the seller depends on two sufficient statistics. First the *total* Poisson intensity of inquiries, $\lambda = \lambda_e + \lambda_v$, and, second, the CDF over the *lowest* markdown conditional on an inquiry,

$$F(m) = \frac{\lambda_e}{\lambda_e + \lambda_v} \sum_{j=0}^{\infty} q_j \left[1 - (1 - G_e(m))^j\right] + \frac{\lambda_v}{\lambda_e + \lambda_v} G_v(m).$$

The first term in this equation is the probability of making an inquiry on the electronic market, multiplied by the probability that the smallest markdown among j offers is less than m. The second term has the same interpretation, but for the voice market.

Given this notation, the Hamilton Jacobi Bellman (HJB) equation for the seller's value at any time $t > 0$ is

$$rV = r - c + \lambda \int \max\{1 - m - V, 0\}\, dF(m) + \gamma(1 - V).$$

The first term on the right-hand side, $r - c$, is the flow value of holding the asset. The second term is the option value of search: the seller makes an inquiry with intensity λ, her best offer is distributed according to $F(m)$, and she accepts if the price $1 - m$ is larger than the value of continuing search, V. The third and last term is the expected flow utility if the seller recovers and exits. As is standard, the HJB shows that the optimal trading strategy of the seller is entirely characterized by a reservation markdown $m^\star \equiv 1 - V$, such that the seller trades if and only if the lowest markdown she receives is less than $m^\star$. Substituting $m^\star = 1 - V$ into the HJB and solving, we obtain our version of McCall's celebrated equation,

$$m^\star = \frac{c}{r+\gamma} - \frac{\lambda}{r+\gamma}\int_0^{m^\star} F(m)\,dm. \tag{1}$$

Appendix A.1 discusses the comparative static of $m^\star$ with respect to parameters. We use this simple model as an aid to interpret our child order data. Recall the child order tree of Figure 1, where a child order is viewed as a sequence of events. Our model implies a probability distribution over this sequence. Namely, there is a new event in the child order tree with intensity $\lambda_e + \lambda_v G_v(m^\star) + \gamma$. Conditional on an arrival, the new event is drawn independently from the arrival time according to the following distribution. The new event is an inquiry without trade on the electronic market with probability

$$\pi_1 = \frac{\lambda_e \sum_{j=0}^{\infty} q_j (1 - G_e(m^\star))^j}{\lambda_e + \lambda_v G_v(m^\star) + \gamma},$$

it is an inquiry with trade on the electronic market with probability

$$\pi_2 = \frac{\lambda_e \sum_{j=0}^{\infty} q_j \left[1 - (1 - G_e(m^\star))^j\right]}{\lambda_e + \lambda_v G_v(m^\star) + \gamma},$$

it is a trade on the voice market with probability

$$\pi_3 = \frac{\lambda_v G_v(m^\star)}{\lambda_e + \lambda_v G_v(m^\star) + \gamma},$$

and it is an exit with probability $\pi_4 = 1 - \pi_1 - \pi_2 - \pi_3$. The formulae above illustrate two sources of bias that make interpreting child order statistics difficult. We discuss these below.

Competing risk bias. First, since the event type is drawn independently from the event arrival time, it follows that the *observed* expected arrival time of any of the four events is given by

$$\overline{\tau} = \frac{1}{\lambda_e + \lambda_v G_v(m^\star) + \gamma}.$$

Notice that this observed expected arrival time is *lower* than the actual arrival time of the event. For example, the actual arrival time of a voice trade is $1/(\lambda_v G(m^\star))$. This is a classical survivor bias or competing risk bias (e.g., Flinn and Heckman, 1982; Katz and Meyer, 1990; Honoré and Lleras-Muney, 2006) created by the arrival of other events. Imagine for example, that sellers exit the market very fast. Then the only trades on the voice market we would observe are those that occur sufficiently quickly, before an exit.

The formulae above show that there is a simple way to correct for this survivor bias: one needs to divide the observed arrival time by the probability of the corresponding event. For example, the true expected time to trade on voice is equal to the ratio $\overline{\tau}/\pi_3$. As we will show below, this correction can be made more generally using a Maximum Likelihood approach, conditional on observable child-order characteristics.

Selection bias. In the data, we can control for observable characteristics of child orders, such as bond type, trade size, and measures of customer connectedness. But there are other characteristics that are difficult to control for based on observables, including the distress cost of a seller, c; her inquiry intensities, λ_e or λ_v; her ability to elicit responses from dealers, $\{q_j\}$; or her exit intensity, γ. Such unobserved characteristics create classical selection issues that could explain the apparent dependence of event probabilities on the number of failed inquiries, shown in Table 6.

To fix ideas formally, suppose that heterogeneity in child orders can be summarized by a one-dimensional type variable $x \in [\underline{x}, \overline{x}]$ which determines the structural variables λ_e, λ_v, γ, c, r, q_j, G_e, G_v, and so on. Then, as we establish in Appendix A.3, the measure of type-x child orders with $n \geq 1$ failed inquiries, $d\mu(x|n)$, satisfies

$$d\mu(x \mid n) = \pi_1(x)^n \, d\mu(x \mid 0), \quad \text{where} \quad \pi_1(x) \equiv \frac{\lambda_e(x) \left(\sum_j q_j \left[1 - G_e(m^\star(x) \mid x)\right]^j \right)}{\lambda_e(x) + \lambda_v(x) G_v(m^\star(x) \mid x) + \gamma(x)}.$$

Hence, the measure of type-x child orders with n failed inquiries declines geometrically with n according to the coefficient, $\pi_1(x)$, which is simply the probability that a type-x inquiry on the electronic trading platform fails to trade (the left-most branch of event 2 in the child-order tree of Figure 1). A consequence of this result is that, as the number of failed inquiries increases, the sample of child orders becomes more selected towards those investors who fail inquiries on the trading platform with higher probability. Hence, if x is unobservable to the econometrician and is monotonically related to x, any outcome variable which is also monotonically related to x will appear to be monotonically related to the number of failed inquiries.

As a concrete example, suppose child orders only differ in terms of the customers' distress cost, c, but are otherwise identical. Then $\pi_1(c)$ is decreasing in c since more distressed sellers have a higher reservation markdown, $m^\star$. As a result, as the number of failed inquiries increases, the sample gets more and more selected towards less distressed customers. It follows that we should observe two key outcome variables—the trading probability and the transaction markdown—decline with the number of failed inquiries, n.

3.2 Evidence about time to trade

In this section, we propose a statistical framework to measure the time it takes customers to trade after their first inquiry on MKTX, correcting for the competing risk bias discussed above, and controlling for observable trade characteristics. Our unit of observation i is an event node in the child order tree of Figure 1—specifically, the type and time of the event that follows an unsuccessful inquiry on MKTX. We index the $K = 4$ possible events by $k \in \{1, \ldots, K\}$, where event $k = 1$ is an inquiry on MKTX without trade, $k = 2$ is an inquiry on MKTX with trade, $k = 3$ is a voice trade, and $k = 4$ is an event that ends the child order or "exit." We assume further that these events arrive at independent exponential times with intensity $\lambda(\theta'_k x_i) = \exp(\theta'_k x_i)$, where x_i is a vector of covariates for that child order. These covariates include trade size, bond characteristics, customers' characteristics, *and* the number of failed inquiries on MKTX; the latter is particularly important, in that it allows us to identify potential duration dependence.

3.2.1 Maximum Likelihood Estimation

Given this framework, conditional on x_i, the event k occurs at time $\tau_i = t$ with probability density

$$\mathbb{P}\left(\tau_i = t, \omega_i = k \mid x_i\right) = \lambda(\theta'_k x_i) e^{-\sum_\ell \lambda(\theta'_\ell x_i)t}.$$

This formula is the product of the probability that event k occurs at time t, $\lambda(\theta'_k x_i)e^{-\lambda(\theta'_k x_i)t}$, and the probability that all other events, $\ell \neq k$, occur *after* time t, $e^{-\sum_{\ell \neq k} \lambda(\theta'_\ell x_i)t}$. This is the sense in which there are competing risks: the probability density accounts for the fact that we observe event k only if the other events $\ell \neq k$ have not occurred before. Aggregating across events and the number of inquiries, the likelihood function is, evidently:

$$\prod_{i=1}^{n}\left(\sum_{k}\mathbb{I}_{\{\omega_i=k\}}\lambda(\theta_k' x_i)e^{-\sum_\ell \lambda(\theta_\ell' x_i)\tau_i}\right).$$

Recall that we only observe whether an exit occurred, and not the time of an exit. Therefore, integrating with respect to τ_i when $\omega_i = K$, we obtain the likelihood for our actual observations:

$$\prod_{i=1}^{n}\left(\sum_{k\neq K}\mathbb{I}_{\{\omega_i=k\}}\lambda(\theta_k' x_i)e^{-\sum_\ell \lambda(\theta_\ell' x_i)\tau_i}+\mathbb{I}_{\{\omega_i=K\}}\frac{\lambda_K(\theta_K' x_i)}{\sum_\ell \lambda(\theta_\ell' x_i)}\right).$$

Taking logs and simplifying, we obtain that the log-likelihood is $\sum_i L(\omega_i, \tau_i, x_i, \theta)$, where:

$$L_i(\omega_i,\tau_i,x_i,\theta)=\sum_{k}\mathbb{I}_{\{\omega_i=k\}}\theta_k' x_i-\mathbb{I}_{\{\omega_i\neq K\}}\left(\sum_{\ell}\exp(\theta_\ell' x_i)\right)\tau_i-\mathbb{I}_{\{\omega_i=K\}}\log\left(\sum_{\ell}\exp(\theta_\ell' x_i)\right).$$

We first gain some qualitative and quantitative intuition by deriving the *unconditional* Maximum Likelihood Estimator (MLE), i.e., the special case in which the only control is a constant.

Lemma 1 *Suppose the only control is a constant, that is, $x_i = 1$ for all observations. Let $\hat{\pi}_k$ denote the empirical frequency of event k and $\hat{\tau}$ the empirical average inter-arrival time of an event $k \neq K$. Then, the MLE of θ_k is $\hat{\theta}_k = \log(\hat{\pi}_k/\hat{\tau})$.*

This is the same estimate that we intuitively derived in the previous section, when discussing the competing risk bias. Indeed, after a failed inquiry, the expected arrival time of any event is $\overline{\tau} = 1/(\sum_\ell \lambda_\ell)$, and the probability of event k is $\pi_k = \lambda_k/\sum_\ell \lambda_\ell$. This shows that $\lambda_k = \pi_k/\overline{\tau}$ and $\theta_k = \log(\lambda_k)$, which is the population counterpart of the estimator in Lemma 1.

The estimation results, shown in Table 7, offer some guidance about the orders of magnitude of arrival times for different events. For example, the unconditional intensity of a voice trade is $e^{-3.367} = 0.0345$ per business hour, corresponding to an average time of $1/0.0345 = 28.99$ business hours, or about 3.3 business days (assuming 9 hours of trading per day). Importantly, the estimates clearly show that competing risk creates a significant bias in calculating time to trade, as 3.22 business days is much larger than the observed average inter-arrival times shown in Table IA.3. Similarly, after a failed inquiry, without controlling for trade and customer characteristics, the time to trade (on MKTX or voice) is $1/\left(e^{-4.077} + e^{-3.367}\right) = 19.43$ business hours, or approximately 2.15 business days.

Next, we move to the *conditional* MLE, with controls for trade characteristics (coefficients shown in Table 8) and for the number of failed inquiries in the child order to date (coefficients shown in Table 9). All controls are dummies. The "baseline" category, defined above, is when all indicator variables are zero: an odd-lot purchase of an investment-grade bond with low time-to-maturity, high turnover, and high amount outstanding during normal times (i.e., not March 2020) for a connected investor. There is no closed form solution for the estimators. However, since the likelihood function is concave in the vector of coefficients $\theta = (\theta_k)_{1\le k\le K}$, it can be maximized reliably using existing optimization packages.

Table 8 shows how the intensities of each event, $\lambda(\theta_k' x)$, vary with trade characteristics. The intensities for the baseline category are obtained by taking the exponential of the intercept. The marginal effect of other trade characteristics is given by the exponential of their respective coefficient. In particular, when the coefficient is sufficiently small, it approximates the marginal effect in percentage term: e.g., from the fourth row in column (2) of Table 8, the intensity of trade with MKTX for a bond rated Ca to C is approximately $-(e^{-0.284} - 1) \simeq 25\%$ lower than for an investment-grade bond.

The estimates in Table 8 demonstrate that intensities vary significantly with trade characteristics. Consider, for example, trade size. We observe that the intensity of trade with MKTX for micro size trades (with size < \$100,000) is larger than for odd lot trades (our baseline category with size between \$100,000 and \$1 million). The intensity for odd lots is larger than for round lots (with size between \$1 and \$5 million), which is larger than for block trades (with size larger than \$5 million). Interestingly the intensity of voice trade is not monotonic in trade size: for example, block trades trade faster on voice. Bonds with low turnover, and high-yield bonds, also have lower trading intensity, both on MKTX and the voice market.

Like the trading probabilities reported earlier, sales and purchases are asymmetric: customers trade faster when they sell, on average, than when they buy. As we discuss in greater detail below, there are several possible reasons for this. For example, it could reflect underlying frictions in locating or "sourcing" the bond, since dealers need to find the bond in order to sell it to a customer but do not to buy it from a customer. Alternatively, it could reflect unobserved heterogeneity, if customer-sellers are, on average, more desperate to trade than customer-buyers.

The last rows of Table 8 show that our measure of customer connectedness is associated with significant heterogeneity in trading intensity on MKTX. Our estimates also indicate that the effects are non-monotonic, which can be consistent with theory. In particular, holding a customer's reservation markdown constant, an increase in connectivity implies that the customer will receive more (or better) offers, so she is more likely to obtain one that falls below her reservation markdown and trade. However, since she expects to receive more offers in the future, her reservation markdown falls and reduces the probability of trading.

Finally, the fifth row of Table 8 reveals that the COVID-19 crisis (identified by inquiries submitted in March 2020) had a significant negative impact on the trading intensity. This finding confirms that market liquidity can deteriorate *along multiple dimensions* in times of stress. In this sense, looking only at the large increases in spreads (documented by, e.g., O'Hara and Zhou, 2021; Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga, 2021, and others) *underestimates* the true affects of sudden selling pressure on market quality.

Table 9 shows that, after controlling for trade characteristics, the number of failed inquiries retains predictive power for the intensity of each event. Interestingly, the intensity of an inquiry on MKTX that doesn't result in trade increases with the number of failed inquiries, but the intensity of successful inquiries—i.e., inquiries on either MKTX or via voice that result in trade—declines. Viewed through the lens of the McCall (1970) model outlined in the previous section, this evidence suggests a role for unobserved child order characteristics. For example, if some types of orders tend to receive few replies, a customer may be forced to make many (frequent) inquiries, knowing that each inquiry is unlikely to generate a good offer. Hence, the composition of child orders with many inquiries could skew towards these types of orders.

3.2.2 Time to trade

We define time to trade as the expected time a customer takes to trade, either on MKTX or on voice, if she is not subject to an exit shock. If the intensities did not depend on the number of failed inquiries, calculating time to trade would be simple. For example, from the intercepts in columns (2) and (3) in Table 8 or 9, the time to trade for our baseline category would be $1/(e^{-3.49} + e^{-3.25}) \simeq 14.43$ business hours, or 1.6 business days. However, the dependence of intensities on the number of failed inquiries requires us to modify this simple formula.

Formally, consider a child order after n failed inquiries. With a slight abuse of notation, let x_n denote the corresponding vector of covariates, where n stands for the number of failed inquiries to date. Then, the expected time to trade satisfies the following recursive formula:

$$
\begin{aligned}
T(x_n) = \mathbb{E}[\tau' \mid x_n] + \mathbb{P}[\omega' = 1 \mid x_n] \times T(x_{n+1}) + \mathbb{P}[\omega' = 2 \mid x_n] \times 0 \\
+ \mathbb{P}[\omega' = 3 \mid x_n] \times 0 + \mathbb{P}[\omega' = 4 \mid x_n] \times T(x_n).
\end{aligned}
$$

The first term is the expected time to the next event. The other terms add up to the expected continuation time to trade after the next event. Specifically, if the next event is an unsuccessful inquiry on MKTX, $\omega' = 1$, then there is one additional failed inquiry and the continuation time to trade is $T(x_{n+1})$. If the next event is $\omega' = 2$ or $\omega' = 3$, then trade occurs so the continuation time to trade is zero. The last line corrects the bias induced by the competing risk of exit: specifically, if the next event is an exit ($\omega' = 4$), we assume that the investor continues to search for a trade instead of exiting, so the continuation time to trade is $T(x_n)$.

Using the exponential formula for expected inter-arrival time and event probability, we obtain

$$
T(x_n) = \frac{1}{\lambda(\theta_1' x_n) + \lambda(\theta_2' x_n) + \lambda(\theta_3' x_n)} + \frac{\lambda(\theta_1' x_n)}{\lambda(\theta_1' x_n) + \lambda(\theta_2' x_n) + \lambda(\theta_3' x_n)} T(x_{n+1}). \tag{2}
$$

We can use this formula to calculate the time to trade. Moreover, differentiating (2) with respect to x, we obtain a corresponding recursive formula for the gradient of time to trade, which allows us to apply the Delta method and obtain standard errors for the time to trade estimates. We illustrate our results in a sequence of figures, where we plot the expected time to trade, conditional on the number of failed inquiries and specific trade characteristics using estimates from the MLE. We represent the 95% confidence intervals by shaded areas surrounding the conditional expectation.

Figure 3 shows that, for our baseline category, the time to trade increases from about two trading days after one failed inquiry to nearly four trading days after ten failed inquiries. High-yield bonds, older bonds, and low turnover bonds have a longer time to trade, though the difference is small relative to other covariates.

In Figure 4, we study the impact of trade size on time to trade. We observe that smaller trades are faster on MKTX. For example, after one failed inquiry, it takes 1.5 days to trade a micro-size bond, while the time it takes to trade a block-size inquiry is almost twice as long. This evidence complements prior studies showing that electronic trading is concentrated on smaller trades (e.g., Hendershott and Madhavan, 2015; O'Hara and Zhou, 2021).

Figure 5 shows that less connected customers, classified as customers who receive fewer offers

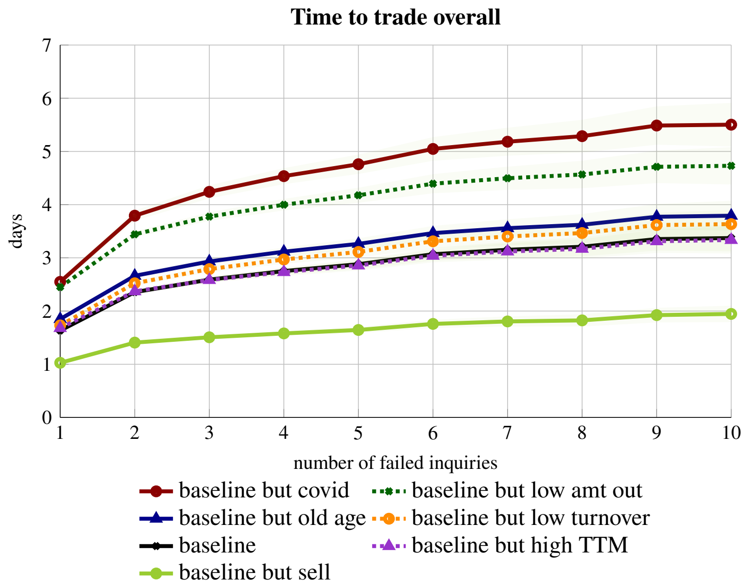


Figure 3. Estimated conditional time to trade from the MLE: observed trade characteristics
This figure plots the estimated time to trade from Equation (2), conditional on the number of failed inquiries and on observed trade characteristics except trade size and customer connectedness categories. "Sell" takes the value of 1 for a sale request, and zero otherwise; "COVID" takes the value of 1 if the RFQ is submitted in March 2020, and zero otherwise; "old age" takes the value of 1 if the bond's age is above the 75th percentile of the distribution, and zero otherwise; "low turnover" takes the value of 1 if the bond's quarterly turnover is below median, and zero otherwise; "high TTM" takes the value of 1 if the bond's time to maturity is above the sample median, and zero otherwise; "low amt out" takes the value of 1 if the bond's amount outstanding is below the sample median, and zero otherwise. The baseline category is an odd-lot purchase of an investment-grade bond, with high turnover, during normal times, for a connected investor, after one failed inquiry.

from dealers, trade much slower on MKTX. For example, in the baseline category, the most connected customers (in the tenth decile of connectedness) trade after approximately 2.2 days following two failed inquiries. For the least connected customers, in deciles 1 to 7, it takes almost 3 times longer to trade.

In Figure 6, we compare time to trade on MKTX to the one on voice for different trade size categories. The first takeaway is that, except for block trades, child orders trade much faster on MKTX than voice. This finding may be explained by the fact that customers initiate their first inquiries on MKTX and prefer to trade on the electronic platform, possibly for its execution quality rather than price discovery. Next, micro-size trades are faster than odd and round lots in both MKTX and the voice market, but block trades are much slower on MKTX. Again, this is not surprising, since, as mentioned above, smaller trades are more likely to be traded on electronic platforms.

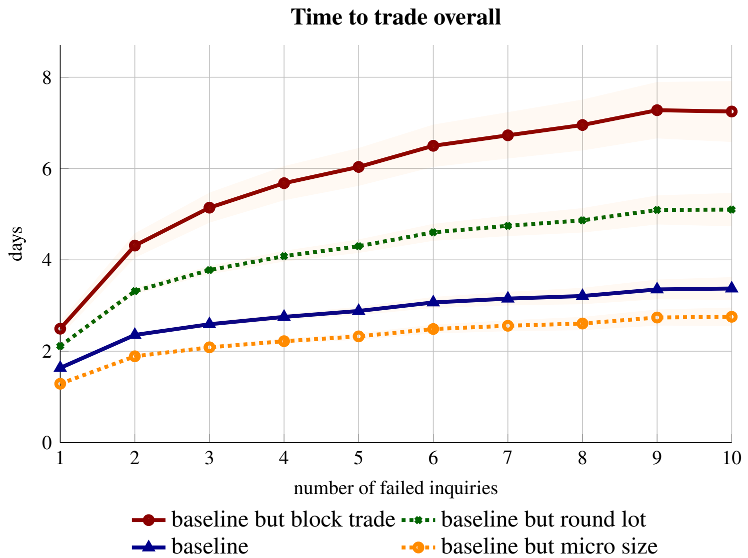


Figure 4. Estimated conditional time to trade from the MLE: impact of size
This figure plots the estimated time to trade from Equation (2), conditional on the number of failed inquiries and on trade size categories, and controlling for other observed trade characteristics. "Micro size" takes the value of 1 if the quantity of dealer response is below \$100,000, and zero otherwise; "odd lot" takes the value of 1 if the quantity of dealer response is between \$100,000 and \$1 million, and zero otherwise; "round lot" takes the value of 1 if the quantity of dealer response is between \$1 million and \$5 million, and zero otherwise; "block trade" takes the value of 1 if the quantity of dealer response exceeds \$5 million, and zero otherwise. The baseline category is an odd-lot purchase of an investment-grade bond, with high turnover, during normal times, for a connected investor, after one failed inquiry.

4 Sources of trading delays

The workhorse model of Duffie, Gârleanu, and Pedersen (2005) posits that trading in OTC markets takes time because investors have to search for a dealer. Hence, if one takes this model literally, a platform like MKTX should eliminate trading delays, as it allows investors to contact dealers instantaneously. And yet, despite the ability to contact multiple dealers at the push of a button, our results show that it can still take investors considerable time to trade in the OTC corporate bond market. How can this be? How can something resembling "search" arise when there are no physical barriers to contacting a counterparty? Our answer is that search models of OTC markets should *not* be taken literally: the assumption that it takes time to find a suitable trade is not meant to capture the time it takes to dial a telephone or push a button. Instead, as Stigler (1961), Demsetz (1968), Pissarides (2000), and others have argued, these models are intended to capture the idea that there are various types of frictions—including information, spatial, and coordination

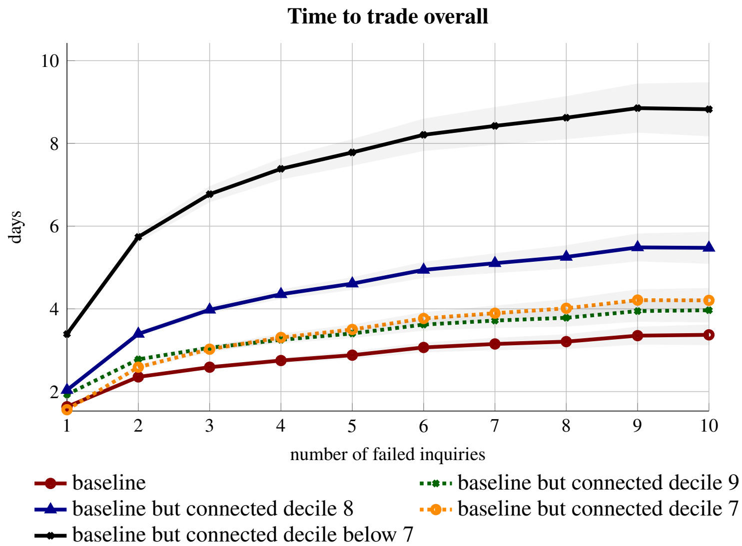


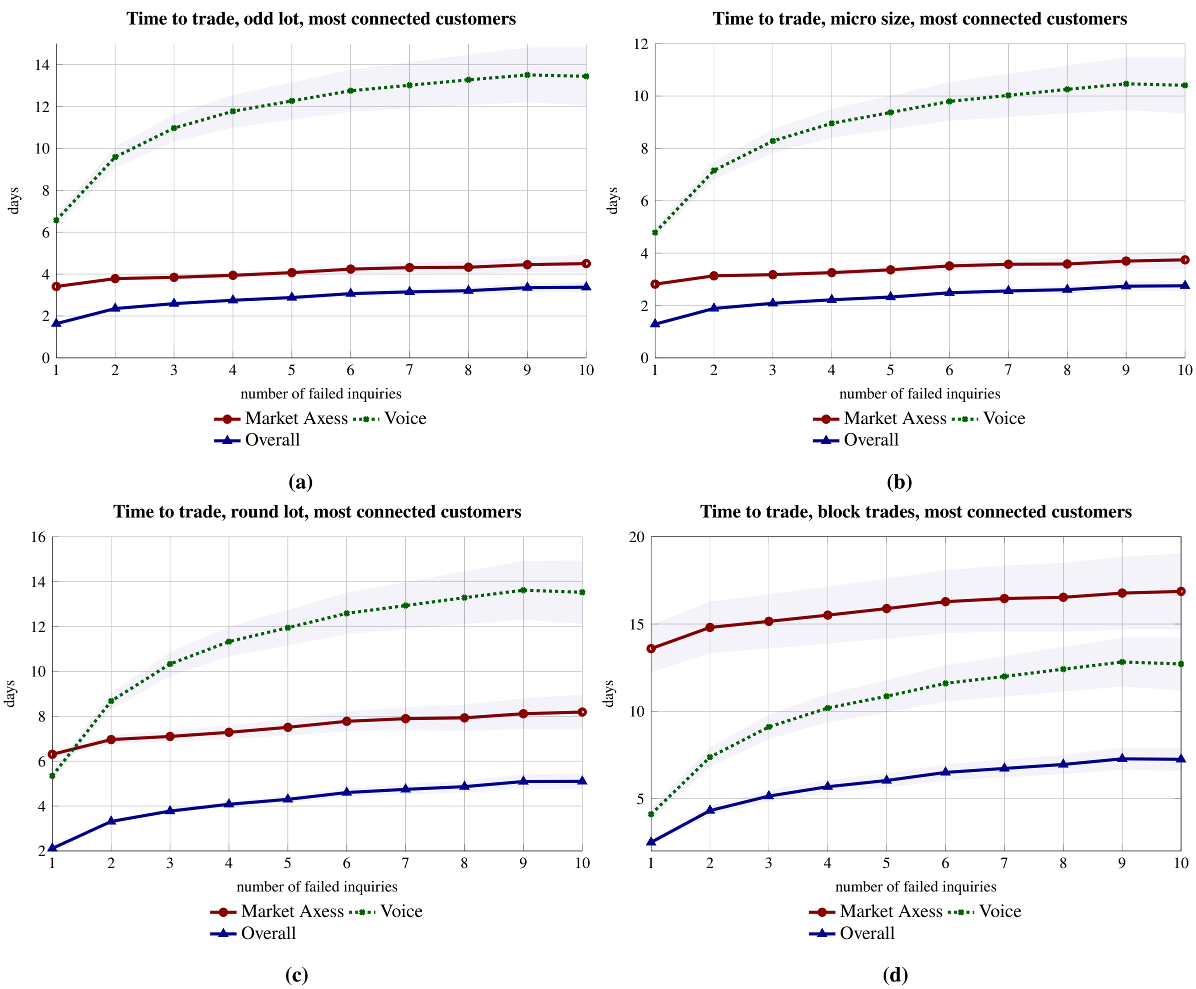
Figure 5. Estimated conditional time to trade from the MLE: impact of customer connectedness
This figure plots the estimated time to trade from Equation (2), conditional on the number of failed inquiries and on customer connectedness categories, and controlling for other observed trade characteristics. We first regress the average number of dealer responses elicited by a particular customer, on that customer's average inquiry size and fractions of requests for sell trades and high-yield bonds. We then rank customers into deciles based on residuals of this regression. "Connected decile 9" is an indicator for the customer being in decile 9, and similarly for other "Connected" indicators. The baseline category is an odd-lot purchase of an investment-grade bond, with high turnover, during normal times, for a connected investor (in decile 10), after one failed inquiry.

frictions—that ultimately result in customers making multiple inquiries over the course of several hours or days before a successful trade occurs. In this section, we dig deeper into the data to shed light on several likely sources of these trading delays, and to provide guidance for future attempts to develop micro-foundations for the reduced-form arrival rates we estimated above.

4.1 Searching for a better quote

In 90% of failed inquiries, customers receive at least one (and often more) quotes from dealers on MKTX. Why do they reject these offers and continue to search? Since bonds with the same CUSIP are homogeneous products, the obvious answer is that customers continue searching in hopes of finding a better quote.

To confirm this basic hypothesis, we construct a measure of the "quality" of replies that a customer receives in response to an inquiry, where quality is defined as the percent deviation of the



(a) (b) (c) (d)

Figure 6. Estimated conditional time to trade from the MLE: MKTX vs. voice
This figure compares the estimated time to trade from Equation (2), conditional on the number of failed inquiries in MKTX vs. voice for the baseline (the top left panel), and different size categories. The baseline category is an odd-lot purchase of an investment-grade bond, with high turnover, during normal times, for a connected investor (in decile 10), after one failed inquiry. Size and connectedness categories are defined in Tables 5 and 8.

best spread in that inquiry relative to the CP+ spread.[18] Specifically, for every child order with at

[18] In the definition of quote quality in Equation (3), "CP+ spread" is defined as (CP+ Ask − CP+ Bid)/(2×CP+ mid), where "CP+ mid" is the midpoint of the CP+ bid and ask, i.e., (CP+ Ask + CP+ Bid)/2. Moreover, the best spread of an inquiry for "buy" inquiries is calculated as $\min_d\{(\text{dealer}_d \text{ quoted price} - \text{CP+ Mid})/\text{CP+ Mid}\}$, and the best spread for an inquiry for "sell" inquiries is calculated as $\min_d\{(\text{CP+ Mid} - \text{dealer}_d \text{ quoted price})/\text{CP+ Mid}\}$.

least two inquiries, we define:

$$\text{quote quality} = \frac{\text{CP+ spread} - \text{best spread for that inquiry}}{\text{CP+ spread}}. \tag{3}$$

Table 10 reports the distribution of quote quality across customers' first inquiries and the relationship between the quality of the best quote and the probability that the customer accepts. As expected, customers are more likely to accept when the best offer is a good one, and much more likely to reject (and, often, continue searching) when the best offer is in the lower deciles of the quote quality distribution. Specifically, we find that customers trade 90% of the time when the quality of their first inquiry is in the top decile, and trade only about 36% when the quote quality is in the bottom decile of the distribution. Of course, average quote quality can differ systematically across bond and trade characteristics, as well as over time. In Table IA.4 in the Internet Appendix, we confirm the positive relationship between quote quality and the probability of acceptance, controlling for a variety of systematic factors.

While these results suggest that customers are more likely to reject low-quality offers, it remains to be shown that continuing to search yields *better* offers. To do so, we construct several statistics intended to measure the extent to which quotes improve (or deteriorate) over the course of a customer's search. To start, we consider the sample of all child orders with (exactly) two inquiries and calculate the difference between the best spread in the second inquiry and the best spread in the first inquiry. Since smaller spreads are "better," a negative value indicates that the best spread improved. Figure 7 plots the distribution of values. We find that the probability that the best offer improved at the second inquiry, relative to the first inquiry, is 0.61.

An alternative definition of spread improvement restricts the sample to child orders that fail at the first inquiry but ultimately result in trade on MKTX. For each child order in this sample, we calculate the difference between the traded spread (relative to the CP+ mid-point price) and the best offered spread in the first inquiry. We find that 71% exhibit spread improvement, i.e., the fraction of values that are negative is 0.71.

To provide a quantitative estimate of the average improvement that customers find for each additional inquiry—controlling for various trade characteristics—we regress this second measure of spread improvement on the number of inquiries that were made before trade occurred. Table 11 reports that, for each additional inquiry in a traded child order, there is an average spread improvement of approximately 3 bps relative to the first inquiry. To put this value in context,

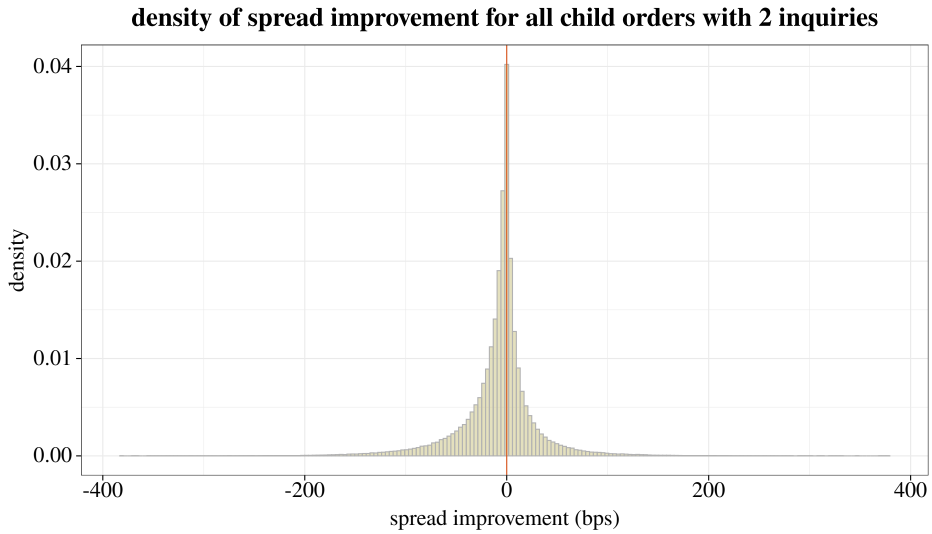


Figure 7. Density of spread improvement
spread improvement is defined as (best spread for the second inquiry − best spread for the first inquiry). The best spread is calculated as $\min_d\{(\text{dealer}_d \text{ quoted price} - \text{CP+ Mid})/\text{CP+ Mid}\}$ for "buy" inquiries, and $\min_d\{(\text{CP+ Mid} - \text{dealer}_d \text{ quoted price})/\text{CP+ Mid}\}$ for "sell" inquiries. The sample consists of all child orders with 2 inquiries.

in our sample, the mean (median) best spread at the inquiry level is 17 bps (10 bps). Hence, this spread improvement is significant, amounting to 18% (30%) of the mean (median) in our data.

4.2 The origins of quote dispersion

We established above that there is considerable heterogeneity in the quality of quotes a customer receives, and that continuing to search can often generate better offers. The next natural question is: What generates dispersion in offers to begin with? A classic explanation, formalized in the theoretical model of Burdett and Judd (1983), is that the number of replies to an inquiry is stochastic.[19]

To explore this theory further, we first construct the probability distribution over the number of replies that an inquiry receives. As one can see in Figure 8, there is considerable dispersion. Of particular importance—from the point of view of the Burdett and Judd (1983) model— is that, in

[19] According to this theory, since dealers do not know how many *other* dealers will respond to an inquiry, they face a trade-off between making a good offer (which will be accepted with high probability, but will yield a low profit) and making a bad offer (which is unlikely to be accepted, but would yield a large profit). In equilibrium, dealers play mixed strategies which generates dispersion.

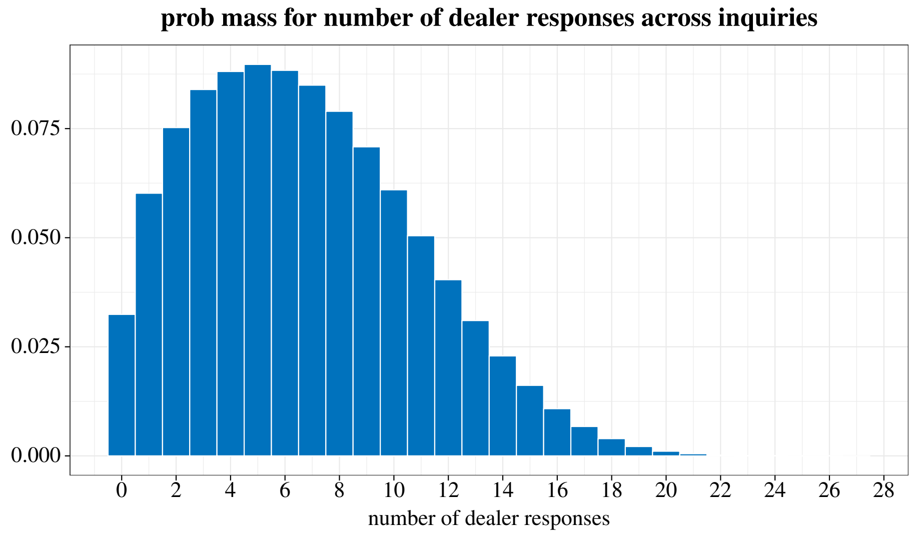


Figure 8. Probability mass function for the number of dealer responses
This figure shows the probability mass for the total number of dealer responses (disclosed and anonymous) at the inquiry level.

the data, the following two probabilities are strictly positive: the probability that an inquiry receives just one offer, and the probability that it receives more than two offers.[20]

From the customers' point of view, additional inquiries can elicit responses from "new" dealers, i.e., dealers that hadn't responded to previous inquiries. In Figure 9, we plot the probability that a new dealer responds to inquiry n, for $n \geq 2$. As one can plainly see, after the first failed inquiry, it is highly likely that a second inquiry will generate a reply from a new dealer. Naturally, this probability declines as the number of failed inquiries increases, albeit relatively slowly.

Finally, linking the two sets of results above, one may wonder whether spread improvement typically arises because a new dealer responds to an inquiry, or because an "incumbent" dealer improves their offer after a failed inquiry. We find that, among the set of traded child orders that exhibit spread improvement, 73% of trades occur with an incumbent dealer (who had replied to a previous inquiry) and 27% occur with a new dealer. Hence, dealers are indeed varying their offer strategy over the course of a child order.

[20]In the model, this is necessary and sufficient for price dispersion. Otherwise, dealers are either monopolist (and set the monopoly price) or competing à la Bertrand (and set the competitive price).

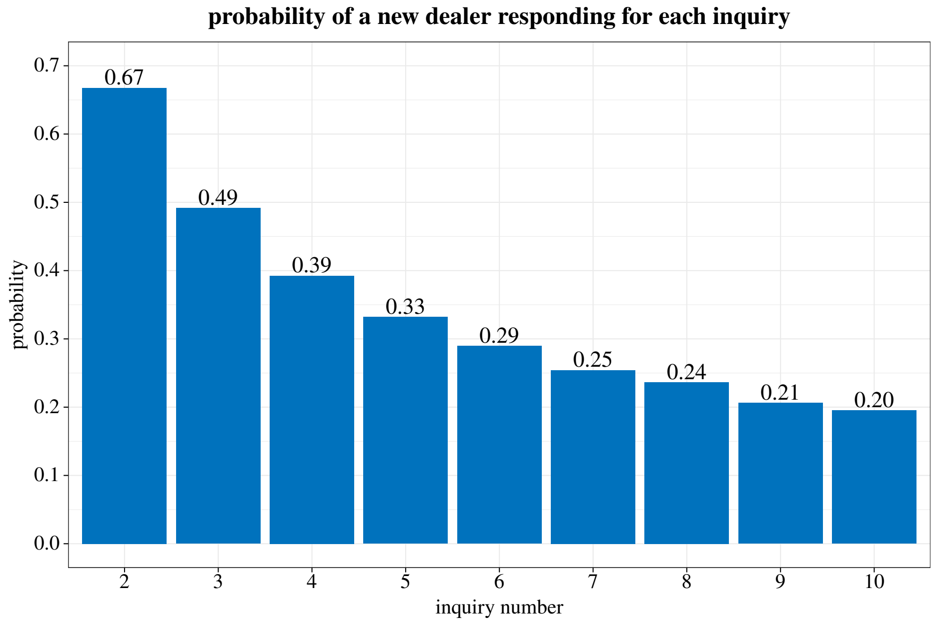


Figure 9. Probability of a response from a new dealer
This figure shows the probability that a new dealer responds to a customer inquiry for a given inquiry number in child orders.

4.3 Dealers' offer strategy

The results above suggest that dealers reply to inquiries probabilistically and—conditional on replying to multiple inquiries within a child order—change their offers over the course of a customer's search. Why do dealers reply to some inquiries in a child order and not others? When they reply to multiple inquiries, do their offers improve over time, deteriorate, or stay the same? We use this final section to explore these questions.

Of course, there are many reasons why a dealer may or may not reply to a particular inquiry. For one, they could simply be busy. Second, a dealer may not be familiar with a particular bond, and thus hesitant to make an offer without time to formulate an informed bid or ask.[21] Lastly, an obvious factor in determining whether a dealer responds to an inquiry—and, in particular, responding to a customer's request to buy—is the dealer's ability to fulfill the order, either with her own inventory or via a quick inter-dealer trade.

[21] Using TRACE data, Cohen (2022) finds that dealers tend to specialize in the bonds that they trade. Moreover, he finds suggestive evidence that specialization is based on information; for example, he finds that dealers are more likely to intermediate two bonds if they are from the same issuer.

The role of inventory. To test the role that a dealer's inventory plays in shaping the likelihood and quality of their reply to a customer inquiry, one would ideally like to directly observe the bonds that a dealer has on its balance sheet and how they change over time. Unfortunately, currently available data does not allow for such direct measures. However, we present below indirect evidence suggesting that the availability of a bond—in particular, whether a dealer has the bond or can acquire it quickly on the inter-dealer market—is an important determinant of whether a dealer replies to an inquiry and the terms of trade she offers.

To start, recall from our estimates in Section 3.2 that customers sell bonds significantly faster than they buy. This asymmetry in trading speed is consistent with the conjecture that inventory is an important determinant of dealers' replies, since owning a particular bond is necessary for a dealer to fulfill a customer-buy inquiry but not a customer-sell inquiry. Digging deeper, we find that the asymmetry between buying and selling extends to the extensive margin as well: for example, the probability that an initial inquiry fails when the customer is trying to make a purchase is significantly higher (32%) than the probability that an initial inquiry fails when the customer is trying to sell (22%).

Indeed, in Table 12, we show that customer-sell inquiries receive significantly more replies, on average, than customer-buy inquiries after controlling for other characteristics of the trade (along with bond, time, customer, and dealer fixed effects). For example, from the first two columns, customer sell requests receive approximately 1.3 more responses from dealers than buy requests. Moreover, the estimates in columns (3) and (4) of Table 12 reveal that, ceteris paribus, inquiries for larger quantities get fewer replies as well. Finally, and perhaps most interestingly, we find that this negative relationship between requested trade size and the number of replies is more pronounced for customer-buys; that is, a customer request to *purchase* a large amount has a more negative effect on the number of replies, relative to a customer request to *sell*.[22] To the best of our knowledge, these facts are new to the literature, and they are also consistent with the conjecture that inventory considerations play an important role in shaping outcomes in the corporate bond market.[23]

To further explore this conjecture, we use the enhanced version of the TRACE data to construct a measure of dealers' *excess inventory* holdings of a particular bond. Intuitively, one might expect

[22]In columns (2) and (3) of Table 12, the regression coefficient for log(Size) gives the marginal effect of increasing size for buy orders, while the sum of the coefficients on log(Size) and log(Size) $\times$ Sell gives the marginal effect of increasing size for sell orders. Hence, if the coefficient on the interaction term is positive, the negative impact of size on number of replies is larger for buy than sell requests.

[23]For a recent theoretical model of OTC trade highlighting the importance of inventory, see Cohen, Kargar, Lester, and Weill (2023).

that dealers' offers to sell a bond would improve when the dealer sector is holding a relatively large amount of the bond, and deteriorate when inventory in the dealer sector shrinks. To test this hypothesis, we define a measure of excess dealer inventory for a given bond as:

$$\text{Excess inventory} = \frac{\text{Current inventory} - \text{Moving avg. of inventory over the last 120 days}}{\text{Moving std. dev. of inventory over the last 120 days}} \tag{4}$$

Tables 13 and 14 report the relationship between the number and the quality of dealers' replies, respectively, and our measure of excess inventory.[24] From Table 13, we see that the number of dealer replies are larger for customer-buy trades and smaller for customer-sell trades when dealer excess inventories are high. For instance, the coefficient from the Poisson regression in column (4) of Table 13 implies that a one unit increase in our measure of excess inventory is associated with a 4.7% ($= e^{0.0462} - 1$) increase in the number of responses to a customer-buy inquiry and approximately a 2.3% ($= e^{0.0462-0.0699} - 1$) decrease in the number of responses to a customer-sell inquiry.

In Table 14, we see that dealers make better offers for customer buy (*sell*) orders when excess inventory is high (*low*), both at the inquiry and child order levels (in columns 1 and 2). For instance, the coefficient from column (4) of Table 14 implies that a one unit increase in our measure of excess inventory is associated with a 2.3 bps decrease in the best offered spread at the inquiry level to a customer-buy inquiry and approximately a 1 bps increase in the best offered spread to a customer-sell inquiry.[25]

Duration dependence. Our estimates of arrival rates in Section 3.2 indicate that time to trade is positively related to the number of failed inquiries in a child order, i.e., that the expected time before a successful trade from the moment inquiry n is placed is less than the expected time to trade starting at the $(n+1)^{th}$ inquiry, for $n \in \mathbb{N}$. One possible explanation of this finding is that dealers alter their behavior over the course of a child order—perhaps they become less likely to reply or the quality of their replies deteriorates. For example, dealers' could change their bidding behavior because of information leakage, as noted by Hendershott and Madhavan (2015), or because they

[24]In Internet Appendix IA.1, we provide a more detailed explanation of how we calculate our measure of dealer excess inventory.

[25]In Figure IA.1 in the Internet Appendix, we show that dealer excess inventories have a corresponding impact on the time-to-trade. We rank inquiries based on a measure of "trade capacity": high trade capacity means either high excess dealer inventory for customer purchases or low dealer excess inventory for customer sales. We show that inquiries with higher trade capacity trade faster, though the economic magnitude of the effect is modest.

learn that the customer was not able to elicit competitive offers, as in the "ringing-phone curse" described in Zhu (2011).

However, an equally plausible alternative is that the relationship between time to trade and the number of failed inquiries derives from unobserved heterogeneity across child orders, and that dealers do not actually change their behavior over the course of a customer's search. For example, if child orders differed according to an unobserved characteristic—including the customers' urgency to trade or their expectations regarding dealers' replies—then they would also differ according to the customers' reservation price, and hence the time to trade.

To tell these two hypotheses apart, we study the dependence of outcome variables on the number of failed inquiries in two ways: controlling for observed trade characteristics and controlling for child-order fixed effects. If unobserved child order characteristics explain the dependence of outcome variables on the number of failed inquiries, then the dependence should disappear after controlling for child order fixed effects. Indeed, when we control for child order fixed effects, we keep *all* child order characteristics fixed, whether they are observed or not.

Table 15 shows the Poisson regression results when the outcome variable is the number of dealer responses. In column (1), we control for observed trade characteristics. We find that holding all observed trade characteristics constant, increasing the number of inquiries from 1 to 2, reduces the number of dealer responses by approximately 27% ($= 1 - e^{-0.311}$). Second, in column (2), we use child order fixed effects instead of trade characteristics. Under this specification, changing the number of inquiries has little impact on the number of dealer responses: in fact, increasing the number of inquiries from 1 to 2 actually *increases* the number of dealer responses by 3.7% ($= 1 - e^{-0.0361}$). Hence, the results in Table 15 provide evidence in favor of the hypothesis that, after controlling for the unobserved characteristics of child orders, the number of replies to an inquiry is largely independent of the number of previous inquiries.

In Table 16, we repeat this regression but for another dependent variable: the spread (trade execution cost) of traded inquiries. As discussed in Section 2, we measure execution cost as a markdown or markup relative to the CP+ benchmark provided by MKTX.[26] The evidence in column (1) suggests that spreads rise as the number of inquiries increases. However, when we control for child order fixed effects, in column (8), we obtain a very different picture: the spreads of

[26]As an alternative measure, we also compute the trading cost measure in Hendershott and Madhavan (2015), which uses the last inter-dealer trade as the reference price for a given bond instead of CP+. Results remain qualitatively similar using this alternative trade execution cost measure.

traded inquiries are much more stable as the number of inquiries within a child order changes and, if anything, go slightly in the opposite direction. To better understand the characteristics of a child order that are being captured by these fixed effects, columns (2) through (7) successively control for customer, bond, issuer, and time fixed effects, along with various interactions. As one can see from column (7), child order fixed effects are not a proxy for simply the customer, the bond, the issuer, or the time when the inquiry was submitted; instead, they appear to capture the state of a particular customer attempting to trade a particular bond at a particular point in time.

These regression results suggest that the trading environment does not change significantly over the course of a child order with multiple inquiries, despite the fact that several outcome variable appear to depend on the number of failed inquiries in the raw data. This finding suggests that either there is minimal learning by customers and/or dealers over the course of a child order search, or that whatever information is revealed has offsetting effects on dealers' offer-making strategies.

5 Conclusion

In this paper, we use data from a leading electronic trading platform to provide new and direct empirical evidence about search frictions in the OTC market for corporate bonds. We start from the observation that when a customer's inquiry on the platform fails to trade, the same customer often returns to the market shortly after to make subsequent inquiries for the same quantity of the same bond. We argue that the resulting sequence of repeated inquiries sheds light on the customers' sequential search process. We estimate that, after a failed inquiry, it takes customers between two and three days to trade. We show that this time to trade depends systematically on trade characteristics and trading venue (electronic vs. voice).

Our analysis brings new insights into the economic mechanisms that lead to trading delays in OTC markets by examining the behavior of customers and dealers throughout the search process. We find that customers who reject an offer and continue to search typically achieve spread improvements. Furthermore, our findings indicate that fluctuations in dealers' inventory holdings are a significant factor contributing to changes in both the quantity and quality of dealer quotes. We also show that customers learning about the distribution of offers contributes to trading delays. Lastly, we provide evidence consistent with unobserved characteristics being a likely reason for the dependence of outcome variables on the number of prior failed attempts to trade.

Overall, our estimates can serve as useful inputs into future quantitative applications of search models while also providing guidance for future theoretical explorations of the micro-foundations of search frictions in OTC markets.

Looking forward, several unique aspects of the MKTX data would allow us to study a number of additional, important questions. For one, since the data allows us to follow (anonymized) customers over time—which is not possible in other commonly used data sets collected from U.S. OTC markets—we can explore how heterogeneity in customers' observable characteristics affects their search behavior and dealers' responses. For example, one could study whether customers who appear "sophisticated" (e.g., in the sense that they trade frequently or have permanent price impact) get better or worse replies from dealers. In the same vein, our ability to track dealers over time would also allow us to better understand the probability and quality of a dealer's reply to an inquiry; again, note that such statistics are simply impossible to derive without observing successful and unsuccessful RFQs.

Since we focused primarily on child orders in the current paper—in order to understand the search process for the most basic unit of trade—a natural next step would be to expand our analysis to understand parent orders more deeply. For example, we could study the incentives of customers to split parent orders into child orders, and the outcomes of child orders *within* a parent order.

Still another source of variation that we did not explore is the difference in outcomes from inquiries that were fulfilled by a customer's "disclosed dealers," who observe their identity, and those fulfilled by others who could not observe the customer's identity. More generally, our data provides fertile ground to test a variety of theoretical results on repeated auctions with imperfect information, a stochastic number of bidders, learning, and so on. We look forward to studying all of these topics in future work.

References

Afonso, Gara, Anna Kovner, and Antoinette Schoar, 2014, Trading partner in the interbank market, Federal Reserve Bank of New York Staff Report no. 620.

Afonso, Gara, and Ricardo Lagos, 2015, Trade dynamics in the market for federal funds, *Econometrica* 83, 263–313.

Bao, Jack, Jun Pan, and Jiang Wang, 2011, The illiquidity of corporate bonds, *Journal of Finance* 66, 911–946.

Bessembinder, Hendrik, Stacey Jacobsen, William Maxwell, and Kumar Venkataraman, 2018, Capital commitment and illiquidity in corporate bonds, *Journal of Finance* 73, 1615–1661.

Bessembinder, Hendrik, William Maxwell, and Kumar Venkataraman, 2006, Market transparency, liquidity externalities, and institutional trading costs in corporate bonds, *Journal of Financial Economics* 82, 251 – 288.

Bessembinder, Hendrik, Chester Spatt, and Kumar Venkataraman, 2020, A survey of the microstructure of fixed-income markets, *Journal of Financial and Quantitative Analysis* 55, 1–45.

Brancaccio, Giulia, and Karam Kang, 2021, Search frictions and product design in the municipal bond market, Working paper, NYU and CMU.

Burdett, Kenneth, and Kenneth L. Judd, 1983, Equilibrium price dispersion, *Econometrica* 51, 955–969.

Cohen, Assa, 2022, Concentration in over-the-counter markets and its impact on financial stability, Working paper, Yeshiva University.

Cohen, Assa, Mahyar Kargar, Benjamin Lester, and Pierre-Olivier Weill, 2023, Inventory, market making, and liquidity: Theory and application to the corporate bond market, Federal Reserve Bank of Philadelphia working paper series.

Demsetz, Harold, 1968, The cost of transacting, *Quarterly Journal of Economics* 82, 33–53.

Deville, Laurent, and Fabrice Riva, 2007, Liquidity and arbitrage in options markets: A survival analysis approach, *Review of Finance* 11, 497–525.

Dick-Nielsen, Jens, 2014, How to clean enhanced TRACE data, Working paper, CBS.

Duffie, Darrell, Nicolae Gârleanu, and Lasse H. Pedersen, 2005, Over-the-counter markets, *Econometrica* 73, 1815–1847.

Edwards, Amy K., Lawrence E. Harris, and Michael S. Piwowar, 2007, Corporate bond market: Transaction costs and transparency, *Journal of Finance* 62, 1421–1451.

Flinn, Christopher, and James Heckman, 1982, New methods for analyzing structural models of labor force dynamics, *Journal of Econometrics* 18, 115–168.

Garbade, Kenneth D., and William L. Silber, 1976, Price dispersion in the government securities market, *Journal of Political Economy* 84, 721–740.

Gavazza, Alessandro, 2016, An empirical equilibrium model of a decentralized asset market, *Econometrica* 84, 1755–1798.

Goldstein, Michael A., Edith S. Hotchkiss, and Erik R. Sirri, 2007, Transparency and liquidity: A controlled experiment on corporate bonds, *Review of Financial Studies* 20, 235–273.

Hendershott, Terrence, Dan Li, Dmitry Livdan, and Norman Schürhoff, 2020, True cost of immediacy, Working paper, UC Berkeley, SFI, and FRB.

Hendershott, Terrence, Dan Li, Dmitry Livdan, Norman Schürhoff, and Kumar Venkataraman, 2021, Quote competition in corporate bonds, Working paper, UC Berkeley, FRB, SFI, and SMU.

Hendershott, Terrence, Dmitry Livdan, and Norman Schürhoff, 2021, All-to-all liquidity in corporate bonds, Working paper, UC Berkeley and SFI.

Hendershott, Terrence, and Ananth Madhavan, 2015, Click or call? auction versus search in the over-the-counter market, *Journal of Finance* 70, 419–447.

Honoré, Bo H., and Adriana Lleras-Muney, 2006, Bounds in competing risks models and the war on cancer, *Econometrica* 74, 1675–1698.

Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill, 2020, Frictional intermediation in over-the-counter markets, *The Review of Economic Studies* 87, 1432–1469.

Kargar, Mahyar, Benjamin Lester, David Lindsay, Shuo Liu, Pierre-Olivier Weill, and Diego Zúñiga, 2021, Corporate bond liquidity during the COVID-19 crisis, *Review of Financial Studies* 34, 5352–5401.

Katz, Lawrence F., and Bruce D. Meyer, 1990, Unemployment insurance, recall expectations, and unemployment outcomes, *Quarterly Journal of Economics* 105, 973–1002.

Kiefer, Nicholas M., and George R. Neumann, 1979, An empirical job-search model, with a test of the constant reservation-wage hypothesis, *Journal of Political Economy* 87, 89–107.

Kondor, Péter, and Gábor Pintér, 2022, Clients' connections: Measuring the role of private information in decentralized markets, *Journal of Finance* 77, 505–544.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1336.

Lo, Andrew W., A. Craig MacKinlay, and June Zhang, 2002, Econometric models of limit-order executions, *Journal of Financial Economics* 65, 31–71.

McCall, John J., 1970, Economics of information and job search, *Quarterly Journal of Economics* 84, 113–126.

O'Hara, Maureen, and Alex Zhou, 2021, The electronic evolution of corporate bond dealers, *Journal of Financial Economics* 140.

Pagnotta, Emiliano S., and Thomas Philippon, 2018, Competing on speed, *Econometrica* 86, 1067–1115.

Pintér, Gábor, and Semih Üslü, 2021, Comparing search and intermediation frictions across markets, Working paper, Bank of England and Johns Hopkins University.

Pissarides, Christopher, 2000, *Equilibrium Unemployment Theory* (MIT Press 2nd Edition, Boston).

Schultz, Paul, 2001, Corporate bond trading costs: A peek behind the curtain, *Journal of Finance* 56, 677–698.

Stigler, George J., 1961, The economics of information, *Journal of Political Economy* 63, 213–225.

Üslü, Semih, 2019, Pricing and liquidity in decentralized asset markets, *Econometrica* 87, 2079–2140.

Weill, Pierre-Olivier, 2020, The search theory of over-the-counter markets, *Annual Review of Economics* 12, 747–773.

Wooldridge, Jeffrey M., 2010, *Econometric Analysis of Cross Section and Panel Data* (MIT Press).

Zhu, Haoxiang, 2011, Finding a good price in opaque over-the-counter markets, *Review of Financial Studies* 25, 1255–1285.

Table 1. Market share of customer trades on MarketAxess

This table presents the market share of MKTX in the US corporate bond for different trade size and credit rating categories. MKTX market shares are calculated as the ratio of trading volume executed on MKTX to the total trading volume observed in TRACE. "Micro size" trades are below $100,000, "Odd lot" trades are between $100,000 and $1 million, "Round lot" trades are between $1 million and $5 million, and "Block trades" exceeds $5 million. Bonds with a rating of Baa3 (on Moody's scale) or better are classified as investment-grade (IG), and those with a Moody's rating of Ba1 or lower are classified as high-yield (HY).

MKTX market share (%)			
	All ratings (1)	IG (2)	HY (3)
All sizes	16.4	20.5	6.4
Micro	26.8	28.9	19.7
Odd lot	46.7	53.8	31.0
Round lot	18.1	26.3	3.8
Block trade	6.6	8.2	0.6

Table 2. Responses of a traded and an untraded inquiry

Panel (a) provides dealers' disclosed responses for a traded inquiry submitted on 08/15/2017 to buy $300,000 of an 11-year, 3.824% investment-grade (USHG) bond issued on 01/17/2017 by Bank of America. The customer received 6 responses, all from dealers, whose anonymized IDs are provided in column (6). Response level (spread over Treasuries for USHG in MKTX) for each dealer response is reported in column (7). In column (10), the response status "Done" flags the response that the submitter accepted, the response status "Cover" flags the second best offer, and the response status "Missed" flags the rest of the responses that the submitter rejected. Panel (b) provides dealer disclosed responses for an untraded inquiry submitted on 08/17/2017 to buy $490,000 of the same bond in panel (a). The customer received 9 responses, all from dealers, whose anonymized IDs and response levels are reported in columns (6) and (7), respectively. The response status "DNT" for this inquiry in column (9) indicates that the inquiry did not trade.

Panel (a): Responses to a traded inquiry on 08/15/2017

Cust. ID (1)	Bond CUSIP (2)	Trade Side (3)	Submit Time (4)	Resp. ID (5)	Dealer ID (6)	Resp. Level (7)	Resp. Quant. (8)	Resp. Status (9)
127	06051GGF0	Buy	08:07:06	1	15420	126.37	300	Missed
127	06051GGF0	Buy	08:07:06	2	16323	129.70	300	Done
127	06051GGF0	Buy	08:07:06	3	11595	128.00	300	Missed
127	06051GGF0	Buy	08:07:06	4	16664	128.05	300	Missed
127	06051GGF0	Buy	08:07:06	5	10392	128.32	300	Missed
127	06051GGF0	Buy	08:07:06	6	12867	128.70	300	Cover

Panel (b): Responses to an untraded inquiry on 08/17/2017

Cust. ID (1)	Bond CUSIP (2)	Trade Side (3)	Submit Time (4)	Resp. ID (5)	Dealer ID (6)	Resp. Level (7)	Resp. Quant. (8)	Resp. Status (9)
127	06051GGF0	Buy	09:56:49	1	15420	125.32	490	DNT
127	06051GGF0	Buy	09:56:49	2	11122	125.70	490	DNT
127	06051GGF0	Buy	09:56:49	3	16377	124.70	490	DNT
127	06051GGF0	Buy	09:56:49	4	12867	125.70	490	DNT
127	06051GGF0	Buy	09:56:49	5	16323	126.20	490	DNT
127	06051GGF0	Buy	09:56:49	6	16664	125.31	490	DNT
127	06051GGF0	Buy	09:56:49	7	10392	125.32	490	DNT
127	06051GGF0	Buy	09:56:49	8	11684	127.01	490	DNT
127	06051GGF0	Buy	09:56:49	9	13910	126.71	490	DNT

Table 3. Cluster of inquiries
This table lists all inquiries by a particular customer (ID 127) for an 11-year, 3.824% investment-grade bond issued on 01/17/2017 by Bank of America over a six-month period in 2017.

Inquiry ID (1)	Cust. ID (2)	Bond CUSIP (3)	Trade Side (4)	Submit Time (5)	Requested Quantity (6)	Inquiry Traded? (7)	Parent Order # (8)	Child Order # (9)
1	127	06051GGF0	Buy	08/15/2017 08:07:06	300	Yes	1	1
2	127	06051GGF0	Buy	08/17/2017 09:56:49	490	No	1	2
3	127	06051GGF0	Buy	08/17/2017 13:57:19	490	Yes	1	2
4	127	06051GGF0	Buy	08/18/2017 08:35:20	290	No	1	3
5	127	06051GGF0	Buy	08/21/2017 08:45:43	290	Yes	1	3
6	127	06051GGF0	Buy	08/23/2017 11:11:38	680	Yes	1	4

Table 4. Parent and child order event statistics
This table reports the fraction of parent and child orders that trade at the first inquiry ("Instant Trade"), the fraction in which we see multiple attempts to trade ("Multiple Attempts"), and the fraction in which we see a single failed inquiry without subsequent match in TRACE ("Abandoned"). A child order is considered to have multiple attempts to trade if it is composed of multiple inquiries on MKTX, or if it has a single failed inquiry that can be matched with a subsequent TRACE record. A parent order is considered to have multiple attempts to trade if it is composed of multiple child orders, or if it has unique child order with multiple trading attempts.

Panel (a): Parent orders

	Instant Trade (1)	Multiple Attempts (2)	Abandoned (3)
Num obs (million)	3.24	1.43	0.79
Fraction total obs	0.59	0.26	0.14
Vol traded ($b of par)	1535	1165	0.00
Fraction of traded vol	0.57	0.43	0.00

Panel (b): Child orders

	Instant Trade (1)	Multiple Attempts (2)	Abandoned (3)
Num obs (million)	5.66	1.03	1.10
Fraction total obs	0.73	0.13	0.14
Vol traded ($b of par)	2,334	366	0.00
Fraction of traded vol	0.86	0.14	0.00

Table 5. Trade probabilities.

This table presents raw trade probabilities at the inquiry and child order levels in columns (1) and (2), and trade probability for child orders at the first inquiry and after failing the first inquiry in columns (3) and (4). "Sell" ("Buy") refers to the subsample of customer sales (purchases); "Investment grade" ("High yield") refers to the subsample of bond that are rated high-grade (high-yield); "Micro size" refers to the subsample in which the quantity of dealer response is below $100,000; "Odd lot" refers to the subsample in which the quantity of dealer response is between $100,000 and $1 million; "Round lot" refers to the subsample in which the quantity of dealer response is between $1 million and $5 million; "Block trade" refers to the subsample in which the quantity of dealer response exceeds $5 million; "High turnover" refers to the subsample in which the bond's quarterly turnover is above median; "High amt outstanding" refers to the subsample in which the bond's amount outstanding is above the sample median; "Old" refers to the subsample in which the bond's age is above the 75th percentile of the distribution. We rank customers into deciles according to the number of dealer responses they receive, after controlling for inquiry size, fraction of requests for sell trades and HY bonds. "Connected" refers to the subsample in which the customer is in deciles 7–10, and "Not connected" refers to the subsample in which the customer is in deciles 1–6.

	Prob. inquiry trades (1)	Prob. child order trades (2)	Prob. child order trades at first inquiry (3)	Prob. child order trades after failed first inquiry (4)
Full sample	0.7060	0.8453	0.7430	0.6261
Sell	0.7519	0.8743	0.7853	0.6576
Buy	0.6591	0.8148	0.6984	0.6012
Investment grade (IG)	0.7323	0.8562	0.7692	0.6149
High yield (HY)	0.5899	0.7948	0.6208	0.6537
Micro size	0.7770	0.9040	0.8082	0.6881
Odd lot	0.6398	0.7943	0.6806	0.5866
Round lot	0.6687	0.7815	0.6921	0.5434
Block trade	0.6936	0.7837	0.7061	0.5011
High turnover	0.7106	0.8505	0.7481	0.6353
Low turnover	0.6917	0.8285	0.7260	0.5951
High amt outstanding	0.7620	0.8831	0.7931	0.6724
Low amt outstanding	0.6434	0.8010	0.6844	0.5928
Old	0.6829	0.8297	0.7208	0.6122
Not old	0.7289	0.8605	0.7645	0.6412
Connected (decile $\geq$ 7)	0.8252	0.9231	0.8456	0.7501
Not connected (decile < 7)	0.4114	0.6392	0.4605	0.4733

Table 6. Child order event statistics

This table presents summary statistics about child order events. A child order can be viewed as a sequence of events, as depicted in Figure 1. Each element of the sequence is one of four possible events: an untraded inquiry on MKTX, a MKTX inquiry with trade, a voice trade, and, if the child order ends without a trade, an exit. By construction, the first event is always either an inquiry on MKTX, without or with trade. The first row shows the probability of a failed and successful inquiry on MKTX. The following rows provides the frequency distribution over the next event in the child order, conditional on the number of failed inquiries to date.

Event	Prob. MKTX inq. w/o trade (1)	Prob. MKTX inq. w trade (2)	Prob. voice trade (3)	Prob. exit (4)
First inquiry	0.2529	0.7430	N/A	N/A
After 1 failed inquiry	0.1604	0.0964	0.2358	0.5075
After 2 failed inquiries	0.3278	0.1066	0.1568	0.4088
After 3 failed inquiries	0.4545	0.1003	0.1185	0.3267
After 4 failed inquiries	0.5469	0.0919	0.0945	0.2667
After 5 failed inquiries	0.6082	0.0850	0.0832	0.2236
After 6 failed inquiries	0.6552	0.0729	0.0691	0.2028
After 7 failed inquiries	0.6925	0.0627	0.0629	0.1819
After 8 failed inquiries	0.7244	0.0672	0.0582	0.1502
After 9 failed inquiries	0.7534	0.0594	0.0493	0.1378
After 10 failed inquiries	0.7624	0.0571	0.0471	0.1335

Table 7. The unconditional Maximum Likelihood Estimator

This table presents estimation results for the unconditional MLE, where the only control is a constant for event $k \in \{1, \ldots, K\}$, with $K = 4$. Event $k = 1$ is an inquiry on MKTX without trade, $k = 2$ is an inquiry on MKTX with trade, $k = 3$ is a voice trade, and $k = 4$ is an exit. Robust standard errors as explained in Chapter 12.5.1 of Wooldridge (2010) are reported in parentheses. Our sample has $N = 1,413,832$ observations.

Event	MKTX inq. w/o trade (1)	MKTX inq. w trade (2)	voice trade (3)	exit (4)
	-3.609^{***}	-4.077^{***}	-3.367^{***}	-2.802^{***}
	(4.80×10^{-6})	(7.47×10^{-6})	(4.51×10^{-6})	(3.11×10^{-6})

Robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 8. The estimated coefficients of the MLE, part 1: trade characteristic dummies
This table presents the first part of our estimation results for the MLE, conditional on trade characteristics (this table) and the number of failed inquiries in the child order to date (in Table 9). "Ba1 to Caa3" takes the value of 1 if the bond's Moody's rating is between Ba1 and Caa3; "Ca to C" is similarly defined; "COVID" takes the value of 1 if the RFQ is submitted in March 2020, and zero otherwise; "High time-to-maturity" takes the value of 1 if the bond's time to maturity is above the sample median, and zero otherwise. We rank customers into deciles according to the number of dealer responses they receive, after controlling for inquiry size, fraction of requests for sell trades and HY bonds. "Connected decile 9" is an indicator for the customer being in decile 9, and similarly for other "Connected" indicators. Other bond and trade characteristics are described in Table 5. Robust standard errors as explained in Chapter 12.5.1 of Wooldridge (2010) are reported in parentheses. Our sample has $N = 1,413,832$ observations.

Event	MKTX inq. w/o trade (1)	MKTX inq. w trade (2)	voice trade (3)	exit (4)
(Intercept)	-4.04^{***}	-3.49^{***}	-3.25^{***}	-2.97^{***}
	(0.0064)	(0.0073)	(0.0058)	(0.0051)
Sell	0.0988^{***}	0.58^{***}	0.241^{***}	-0.0234^{***}
	(0.0041)	(0.0055)	(0.0038)	(0.0034)
Ba1 to Caa3	0.00794^{*}	-0.0505^{***}	0.168^{***}	-0.164^{***}
	(0.0047)	(0.0061)	(0.0043)	(0.0039)
Ca to C	-0.0141	-0.284^{***}	0.427^{***}	-0.182^{***}
	(0.053)	(0.075)	(0.042)	(0.043)
Covid	-0.172^{***}	-0.467^{***}	-0.249^{***}	-0.176^{***}
	(0.01)	(0.013)	(0.0081)	(0.0077)
Old	0.0066^{*}	-0.0957^{***}	-0.0587^{***}	0.0381^{***}
	(0.0041)	(0.0053)	(0.0038)	(0.0034)
Turnover below median	-0.0108^{***}	-0.127^{***}	-0.0758^{***}	0.13^{***}
	(0.0045)	(0.006)	(0.0045)	(0.0038)
High time-to-maturity	0.00553^{*}	0.0419^{***}	-0.0825^{***}	0.0875^{***}
	(0.0041)	(0.0053)	(0.0038)	(0.0034)
Low amt outstanding	0.141^{***}	-0.282^{***}	-0.356^{***}	0.15^{***}
	(0.0042)	(0.0053)	(0.0037)	(0.0034)
Micro size	0.023^{***}	0.187^{***}	0.403^{***}	-0.262^{***}
	(0.0041)	(0.0054)	(0.0039)	(0.0035)
Round lot	-0.159^{***}	-0.445^{***}	-0.0305^{***}	0.363^{***}
	(0.0082)	(0.011)	(0.0081)	(0.0063)
Block trade	-0.415^{***}	-1.26^{***}	0.0611^{**}	0.531^{***}
	(0.031)	(0.044)	(0.029)	(0.023)
Connected decile < 7	-0.0173^{***}	-1.74^{***}	-0.227^{***}	0.0746^{***}
	(0.0055)	(0.008)	(0.0053)	(0.0045)
Connected decile 7	0.00634	-1.1^{***}	0.226^{***}	0.0412^{***}
	(0.0066)	(0.0087)	(0.0063)	(0.0054)
Connected decile 8	0.141^{***}	-0.713^{***}	0.236^{***}	0.0538^{***}
	(0.0068)	(0.0084)	(0.0064)	(0.0057)
Connected decile 9	0.0662^{***}	-0.176^{***}	0.0779^{***}	0.0896^{***}
	(0.0066)	(0.0072)	(0.0063)	(0.0055)
Failed inquiry controls	Yes	Yes	Yes	Yes

Robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 9. The estimated coefficient of the MLE, part 2: the failed inquiries dummies
This table presents the second part of our estimation results for the MLE, conditional on trade characteristics (in Table 8) and the number of failed inquiries in the child order to date (this table). For convenience, we repeat the estimates for the baseline category, i.e., the intercept from Table 8. Event $k = 1$ is an inquiry on MKTX without trade, $k = 2$ is an inquiry on MKTX with trade, $k = 3$ is a voice trade, and $k = 4$ is an exit. "Failed j" takes the value of 1 if the number of failed inquiries in the child order to date is equal to j, and zero otherwise. Robust standard errors as explained in Chapter 12.5.1 of Wooldridge (2010) are reported in parentheses. Our sample has $N = 1,413,832$ observations.

Event	MKTX inq. w/o trade (1)	MKTX inq. w trade (2)	voice trade (3)	exit (4)
(Intercept)	-4.04^{***}	-3.49^{***}	-3.25^{***}	-2.97^{***}
	(0.0064)	(0.0073)	(0.0058)	(0.0051)
Failed 2	0.516^{***}	-0.0574^{***}	-0.546^{***}	-0.342^{***}
	(0.0052)	(0.0072)	(0.0055)	(0.0046)
Failed 3	0.854^{***}	-0.0474^{***}	-0.796^{***}	-0.565^{***}
	(0.0074)	(0.012)	(0.01)	(0.0081)
Failed 4	1.07^{***}	-0.0677^{***}	-0.981^{***}	-0.761^{***}
	(0.0098)	(0.019)	(0.017)	(0.013)
Failed 5	1.19^{***}	-0.0621^{***}	-1.07^{***}	-0.919^{***}
	(0.012)	(0.026)	(0.024)	(0.019)
Failed 6	1.34^{***}	-0.108^{***}	-1.17^{***}	-0.96^{***}
	(0.015)	(0.036)	(0.035)	(0.026)
Failed 7	1.39^{***}	-0.236^{***}	-1.19^{***}	-1.08^{***}
	(0.019)	(0.049)	(0.044)	(0.035)
Failed 8	1.47^{***}	-0.0579	-1.27^{***}	-1.37^{***}
	(0.021)	(0.056)	(0.059)	(0.049)
Failed 9	1.57^{***}	-0.239^{***}	-1.5^{***}	-1.41^{***}
	(0.025)	(0.073)	(0.081)	(0.061)
Failed 10	1.77^{***}	-0.228^{***}	-1.43^{***}	-1.37^{***}
	(0.014)	(0.044)	(0.046)	(0.034)
Trade char. controls	Yes	Yes	Yes	Yes

Robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 10. Quote quality for the first inquiry in child orders

In this table, we sort child orders into deciles by the "quote quality" of the first inquiry, defined in Equation (3), and report average quote quality and trade probabilities. Quote quality is winsorized at the 5% and 95% levels.

Decile	Quote quality	Trade probability
10	3.314	0.900
9	1.543	0.897
8	0.536	0.881
7	-0.202	0.860
6	-0.879	0.831
5	-1.597	0.794
4	-2.464	0.739
3	-3.689	0.660
2	-5.891	0.540
1	-11.319	0.360

Table 11. spread improvement relative of the first inquiry

This table reports the estimates of regressing "spread improvement" on the number of inquiries in child orders. In column (1), we include trade characteristics described in Tables 5 and 8 and year-month fixed effects. In column (2), in addition to indicators for inquiry number in child orders, we add year-month and child order fixed effects to control for unobserved heterogeneity. "spread improvement" is defined as the difference between the spread at the last (traded) inquiry and the best spread offered at the first inquiry, for all traded child orders with > 1 inquires. As discussed in Section 2, we measure execution costs as a markdown or markup relative to the benchmark provided by MKTX, called Composite+. The sample includes all child orders submitted by customers that result in trade, have at least two inquiries, and get a response on their first inquiry. Clustered standard errors at the customer level are shown in parentheses.

Dependent Variable:	spread improvement			
Model:	(1)	(2)	(3)	(4)
Variables				
Number of inquiries	-3.267***	-3.015***	-2.866***	-2.826***
	(0.5626)	(0.5522)	(0.4843)	(0.4822)
Bond & Trade controls	Yes	Yes	Yes	Yes
Customer controls	Yes			
Fixed-effects				
year-month	Yes	Yes	Yes	Yes
customer		Yes	Yes	Yes
bond			Yes	Yes
dealer				Yes
Fit statistics				
Observations	107,429	107,429	107,429	107,429
R^2	0.12129	0.15177	0.28545	0.29322
Within R^2	0.09703	0.07240	0.02850	0.02766

Clustered (customer) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 12. Number of dealer responses: effect of order size and trade direction
This table reports the OLS regression estimates for the impact of trade side (buy vs. sell) and requested quantity (size) on the number of responses received from dealers. Clustered standard errors at the customer level are shown in parentheses.

Dependent Variable:	number of responses			
Model:	(1)	(2)	(3)	(4)
Variables				
Sell	1.281***	1.340***	0.2801***	0.4836***
	(0.0500)	(0.0550)	(0.1085)	(0.1077)
log(Size)			-0.3484***	-0.3430***
			(0.0245)	(0.0275)
Sell × log(Size)			0.1939***	0.1636***
			(0.0233)	(0.0226)
Fixed-effects				
day	Yes	Yes	Yes	Yes
customer	Yes	Yes	Yes	Yes
dealer	Yes	Yes	Yes	Yes
bond		Yes		Yes
Fit statistics				
Observations	8,365,989	8,365,989	8,365,989	8,365,989
R^2	0.40195	0.61298	0.41074	0.62170
Within R^2	0.03654	0.05927	0.05071	0.08047

Clustered (customer) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 13. Number of dealer responses and excess dealer inventories
This table presents Poisson regression estimates for number of dealer responses on the inquiry number in child orders, dealer excess inventory, defined in Equation (4), and its interaction with an indicators for a sell trade. "Inquiry j" takes the value of 1 if it is the jth inquiry in the child order. In columns (1) and (2) we include bond and trade characteristics described in Tables 5 and 8. In columns (3) and (4), we control for the unobserved child order characteristics by adding child order fixed effects to the regression. The sample includes child orders that have at least two inquiries and excludes inquiries submitted by dealers. Clustered standard errors at the customer level are shown in parentheses.

Dependent Variable:	number of responses			
Model:	(1)	(2)	(3)	(4)
Variables				
Inquiry 2	0.0287***	0.0285***	0.0308***	0.0305***
	(0.0040)	(0.0040)	(0.0041)	(0.0041)
Inquiry 3	-0.0139*	-0.0130	0.0526***	0.0519***
	(0.0080)	(0.0080)	(0.0074)	(0.0073)
Inquiry 4	-0.0310**	-0.0297**	0.0618***	0.0606***
	(0.0131)	(0.0130)	(0.0095)	(0.0094)
Inquiry 5	-0.0342*	-0.0328*	0.0723***	0.0707***
	(0.0183)	(0.0182)	(0.0150)	(0.0150)
Inquiry 6	-0.0313	-0.0297	0.0800***	0.0780***
	(0.0191)	(0.0191)	(0.0158)	(0.0157)
Inquiry 7	-0.0331	-0.0317	0.0867***	0.0842***
	(0.0212)	(0.0213)	(0.0157)	(0.0156)
Inquiry 8	-0.0305	-0.0296	0.0905***	0.0874***
	(0.0199)	(0.0199)	(0.0212)	(0.0212)
Inquiry 9	-0.0353	-0.0343	0.0887***	0.0848***
	(0.0247)	(0.0250)	(0.0224)	(0.0226)
Inquiry $\geq$ 10	-0.0340	-0.0334	0.0985***	0.0925***
	(0.0210)	(0.0213)	(0.0246)	(0.0242)
Dealer excess inventory		0.0311***		0.0462***
		(0.0017)		(0.0028)
Dealer excess inventory $\times$ Sell		-0.0547***		-0.0699***
		(0.0037)		(0.0061)
Bond & trade controls	Yes	Yes		
Fixed-effects				
customer	Yes	Yes		
bond	Yes	Yes		
issuer	Yes	Yes		
year-month	Yes	Yes	Yes	Yes
child order			Yes	Yes
Fit statistics				
Observations	968,886	968,601	969,477	969,188
Squared Correlation	0.56289	0.56502	0.90485	0.90501
Pseudo R^2	0.21306	0.21408	0.34781	0.34787
BIC	4,221,108.2	4,214,780.9	8,965,014.4	8,962,772.8

Clustered (customer) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 14. Spread and excess dealer inventories

This table presents estimates for regressing trade execution costs on the inquiry number in child orders, dealer excess inventory, defined in Equation (4), and its interaction with an indicators for a sell trade. "Inquiry j" takes the value of 1 if it is the jth inquiry in the child order. In columns (1) and (2) we include bond and trade characteristics described in Tables 5 and 8. In columns (3) and (4), we control for the unobserved child order characteristics by adding child order fixed effects to the regression. As discussed in Section 2, we measure execution costs as a markdown or markup relative to the benchmark provided by MKTX, called Composite+. The sample includes child orders that have at least two inquiries and excludes inquiries submitted by dealers. Clustered standard errors at the customer level are shown in parentheses.

Dependent Variable:	inquiry best spread			
Model:	(1)	(2)	(3)	(4)
Variables				
Inquiry 2	-5.768***	-5.764***	-5.445***	-5.390***
	(0.5596)	(0.5602)	(0.6624)	(0.6596)
Inquiry 3	-4.470***	-4.517***	-8.081***	-7.936***
	(0.4345)	(0.4387)	(1.130)	(1.125)
Inquiry 4	-4.322***	-4.388***	-9.927***	-9.701***
	(0.2893)	(0.2919)	(1.498)	(1.486)
Inquiry 5	-3.955***	-4.020***	-10.99***	-10.62***
	(0.4951)	(0.4882)	(1.808)	(1.784)
Inquiry 6	-3.465***	-3.538***	-11.20***	-10.83***
	(0.4608)	(0.4471)	(2.075)	(2.056)
Inquiry 7	-2.872***	-2.909***	-11.10***	-10.59***
	(0.6680)	(0.6517)	(2.352)	(2.326)
Inquiry 8	-3.090***	-3.147***	-10.96***	-10.40***
	(0.8697)	(0.8593)	(2.692)	(2.672)
Inquiry 9	-5.135***	-5.201***	-12.99***	-12.40***
	(0.8155)	(0.8160)	(2.852)	(2.836)
Inquiry $\geq$ 10	-3.282***	-3.314***	-12.40***	-11.54***
	(0.9341)	(0.9018)	(3.215)	(3.252)
Dealer excess inventory		-1.646***		-2.281***
		(0.1250)		(0.2723)
Dealer excess inventory $\times$ Sell		2.296***		3.317***
		(0.1927)		(0.2906)
Bond & trade controls	Yes	Yes		
Fixed-effects				
customer	Yes	Yes		
bond	Yes	Yes		
day	Yes	Yes	Yes	Yes
child order			Yes	Yes
Fit statistics				
Observations	902,979	902,738	903,548	903,303
R^2	0.34366	0.34489	0.80506	0.80516
Within R^2	0.09393	0.09565	0.07582	0.07606

Clustered (customer) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 15. Poisson model for the number of dealer responses

This table presents Poisson regression estimates for number of dealer responses on indicators for the inquiry number in child orders. "Inquiry j" takes the value of 1 if it is the jth inquiry in the child order. In column (1) we include trade characteristics described in Tables 5 and 8. In column (2), we control for the unobserved child order characteristics by adding child order fixed effects to the regression. The sample excludes inquiries submitted by dealers.

Dependent Variable:	number of dealer responses	
Model:	(1)	(2)
Variables		
(Intercept)	1.903***	
	(0.0003)	
Inquiry 2	-0.3110***	0.0361***
	(0.0008)	(0.0006)
Inquiry 3	-0.4241***	0.0607***
	(0.0016)	(0.0012)
Inquiry 4	-0.4724***	0.0670***
	(0.0027)	(0.0019)
Inquiry 5	-0.4799***	0.0812***
	(0.0038)	(0.0026)
Inquiry 6	-0.4990***	0.0867***
	(0.0051)	(0.0034)
Inquiry 7	-0.5211***	0.0837***
	(0.0066)	(0.0044)
Inquiry 8	-0.5163***	0.1017***
	(0.0081)	(0.0053)
Inquiry 9	-0.5156***	0.1065***
	(0.0097)	(0.0064)
Inquiry ≥ 10	-0.4973***	0.1055***
	(0.0055)	(0.0069)
Trade char. controls	Yes	
Fixed-effects		
child order		Yes
Fit statistics		
Observations	9,455,325	9,108,063
Squared Correlation	0.33738	0.99172
Pseudo R^2	0.14526	0.36693
BIC	45,117,005.9	165,520,283.0
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*		

Table 16. Unobserved heterogeneity: Trade execution cots
This table reports the estimates of regressing inquiry spreads on indicators for inquiries in child orders. In column (1), we include trade characteristics described in Tables 5 and 8 and year-month fixed effects. In column (2), in addition to indicators for inquiry number in child orders, we add year-month and child order fixed effects to control for unobserved heterogeneity. As discussed in Section 2, we measure execution costs as a markdown or markup relative to the benchmark provided by MKTX, called Composite+. Clustered standard errors at the customer level are shown in parentheses.

Dependent Variable:	best spread							
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
Inquiry 2	6.031***	4.577***	4.645***	4.887***	5.283***	1.764***	-0.5204**	-5.318***
	(0.5350)	(0.5829)	(0.6223)	(0.5601)	(0.4656)	(0.4072)	(0.2515)	(0.6117)
Inquiry 3	8.539***	6.559***	6.518***	7.065***	7.351***	2.298***	-1.097**	-7.867***
	(1.100)	(1.046)	(1.211)	(1.047)	(0.9341)	(0.6119)	(0.4783)	(1.022)
Inquiry 4	8.337***	6.486***	6.198***	7.117***	7.072***	1.827**	-2.067***	-9.640***
	(1.655)	(1.555)	(1.889)	(1.521)	(1.412)	(0.8449)	(0.5912)	(1.392)
Inquiry 5	8.029***	6.448***	6.019***	7.193***	6.894***	1.689*	-2.550***	-10.59***
	(2.056)	(1.818)	(2.318)	(1.782)	(1.778)	(1.003)	(0.8368)	(1.709)
Inquiry 6	7.750***	6.273***	5.678**	7.430***	6.575***	1.634*	-2.969***	-10.82***
	(2.290)	(1.991)	(2.605)	(1.904)	(1.891)	(0.9102)	(1.004)	(2.039)
Inquiry 7	7.018***	5.984***	5.104*	7.289***	5.869***	1.576*	-2.981***	-10.57***
	(2.435)	(2.102)	(2.787)	(1.917)	(2.027)	(0.8794)	(1.091)	(2.353)
Inquiry 8	5.883**	5.312***	4.038	6.947***	5.021***	1.778**	-2.466*	-9.984***
	(2.329)	(1.926)	(2.674)	(1.705)	(1.931)	(0.8509)	(1.268)	(2.687)
Inquiry 9	3.337	2.852	1.523	4.995***	2.756	0.0791	-4.280***	-11.66***
	(2.186)	(1.920)	(2.812)	(1.783)	(1.803)	(0.8365)	(1.275)	(2.869)
Inquiry ≥ 10	4.463**	3.581**	1.528	7.157***	3.820**	1.961***	-3.070**	-10.05***
	(2.104)	(1.818)	(2.810)	(1.696)	(1.772)	(0.7102)	(1.212)	(3.079)
Trade char. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Fixed-effects								
yearmon	Yes	Yes	Yes					Yes
customer		Yes	Yes					
bond		Yes						
issuer			Yes					
customer-yearmon				Yes				
issuer-yearmon					Yes			
customer-issuer-yearmon						Yes		
customer-bond-yearmon							Yes	
child order								Yes
Fit statistics								
Observations	7,244,316	7,244,316	7,244,316	7,244,316	7,244,316	7,244,316	7,244,316	7,244,316
R^2	0.17731	0.25053	0.21075	0.21559	0.23203	0.53163	0.72244	0.95995
Within R^2	0.16438	0.07363	0.08300	0.11253	0.11123	0.06916	0.06167	0.09171

Clustered (customer) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Appendix

A More on the McCall (1970) Model of Section 3.1

A.1 Comparative statics for the reservation markdown

Recall that the reservation markdown of a seller is the unique solution of

$$m^\star = \frac{c}{r+\gamma} - \frac{\lambda}{r+\gamma}\int_0^{m^\star} F(m)\,dm.$$

The first term in Equation (1), $c/(r+\gamma)$, is the expected present value of the seller's distress cost. It represents the monopsony markdown: the maximum markdown a seller would be willing to accept if she received just one take-it-or-leave-it offer by a dealer, and no offer forever after. The optimal reservation markdown is less than the monopsony markdown because of the option value of searching for another offer.

By inspection, one sees that $m^\star$ admits the following comparative statics. It increases with the distress cost c, decreases with the interest rate, r, decreases with the exit rate, γ, decreases with the inquiry intensity, λ, and increases in response to first-order stochastic dominance shift in the distribution of the best markdown, $F(m)$.

These comparative statics are similar to the one obtained in the classical job-search setting except for the one with respect to $r+\gamma$. The reason is that, in our setting, increasing $r+\gamma$ impacts the seller's problem in two ways. First, just as in job-search models, it reduces the option value of search which, all else equal, increases the reservation markdown. Second, and new to this setting, it decreases the present value of seller's distress costs, which decreases the reservation markdown. The second effect, it turns out, always dominates in our setting.

A.2 Alternative specifications of the exit shock

In the text we interpreted the exit shock as recovery from distress. In the data, the exit of a child order could arise for other reasons, in particular because the customer updates the quantity she demands or supplies. In the McCall (1970), this means that the continuation value of exit is not necessarily equal to the par value of the bond, but to some other value which we denote by $1-\hat{m}$.

The HJB equation becomes

$$rV = r - c + \lambda \int \max\{1 - m - V, 0\}\, dF(m) + \gamma(1 - \hat{m} - V).$$

The optimal trading strategy of the customer remains characterized by a reservation markdown characterized by the equation:

$$m^\star = \frac{c + \gamma \hat{m}}{r + \gamma} - \frac{\lambda}{r + \gamma} \int_0^{m^\star} F(m)\, dm.$$

In particular, this shows that the competing risk bias does not depend on the nature of the exit shock. All that matters is that exit censors the sample of successful child orders.

A.3 Heterogeneity

As in the main body of the paper, assume that heterogeneity in child order is summarized by the one dimensional type-variable $x \in [\underline{x}, \overline{x}]$. Assume that, an any point in time, there is an inflow $d\phi(x)$ of type-x child orders in the market. Then, the measure of type-x child orders with $n \geq 0$ failed inquiries satisfies the inflow-outflow equations:

$$\begin{aligned}
n = 0 \;:\; & d\phi(x) = d\mu(x \mid 0) \left(\lambda_e(x) + \lambda_v(x) G_v(m^\star(x) \mid x) + \gamma(x)\right) \\
n \geq 1 \;:\; & \lambda_e(x) \left(\sum_j q_j \left[1 - G_e(m^\star(x) \mid x)\right]^j \right) d\mu(x \mid n-1) \\
& \qquad = d\mu(x \mid n) \left(\lambda_e(x) + \lambda_v(x) G_v(m^\star(x) \mid x) + \gamma(x)\right).
\end{aligned}$$

The left-hand side is the inflow: for example, in the second equation, it is composed of all those customers who make inquiries on the trading platform but fail to trade. Correspondingly, the right-hand side is the outflow: in the second equation, it is composed of all investors who make inquiries on the trading platform, trade on the voice market, or exit. Taken together, these inflow-outflow equations imply:

$$d\mu(x \mid x) = \pi_1(x)^n d\mu(x \mid 0) \text{ where } \pi_1(x) \equiv \frac{\lambda_e(x) \left(\sum_j q_j \left[1 - G_e(m^\star(x) \mid x)\right]^j \right)}{\lambda_e(x) + \lambda_v(x) G_v(m^\star(x) \mid x) + \gamma(x)}, \tag{5}$$

and $d\mu(x \mid 0) = d\phi(x)/\big(\lambda_e(x) + \lambda_v(x)G_v(m^\star(x) \,|\, x) + \gamma(x)\big)$. According to (5), the measure of type-x child orders with n failed inquiries declines with n geometrically. The geometric coefficient is simply the probability of failing an inquiry on MarketAxess, in the child-order tree.

Next we show that the direction of the selection bias depends on the geometric coefficient, $\pi_1(x)$. Namely, let

$$dH(x \mid n) = \frac{d\mu(x \mid n)}{\int_{\underline{x}}^{\overline{x}} d\mu(y \mid n)}$$

denote probability distribution over x conditional on n. We obtain the following Lemma:

Lemma 2 *If $\pi_1(x)$ is an increasing (decreasing) function, then $H(x \mid n)$ first-order stochastically dominates (is first-order stochastically dominated by) $H(x \mid n-1)$.*

Lemma 2 shows that as the number of failed inquiries, n, increases, the sample of child order becomes more selected towards those investors who, in their child order tree, fail inquiries on the trading platform with higher probability. As a result, if x is unobservable to the econometrician, any outcome variable which is monotonically related to x will appear to be monotonically related to the number of failed inquiries.

We prove the Lemma for the case of an increasing $\pi_1(x)$. We start from the definition of H:

$$dH(x \mid n) = \frac{\pi_1(x)\, d\mu(x \mid n-1)}{\int_{\underline{x}}^{\overline{x}} \pi_1(y)\, d\mu(y \mid n-1)} = \frac{\pi_1(x)\, dH(x \mid n-1)}{\int_{\underline{x}}^{\overline{x}} \pi_1(y)\, dH(y \mid n-1)}$$

where the first equality follows from the recursion $d\mu(x \mid n) = \pi_1(x)\, d\mu(x \mid n-1)$, and the second equality follows from dividing both the numerator and the denominator by $\int_{\underline{x}}^{\overline{x}} d\mu(x \mid n-1)$. Therefore:

$$\begin{aligned}
&\operatorname{sign}\left(H(x \mid n) - H(x \mid n-1)\right) \\
&= \operatorname{sign}\left(\frac{\int_{\underline{x}}^{x} \pi_1(y)\, dH(y \mid n-1)}{\int_{\underline{x}}^{\overline{x}} \pi_1(z)\, dH(z \mid n-1)} - \int_{\underline{x}}^{x} dH(y \mid n-1)\right) \\
&= \operatorname{sign}\left(\int_{\underline{x}}^{x} \left[\pi_1(y) - \int_{\underline{x}}^{\overline{x}} \pi_1(z)\, dH(z \mid n-1)\right] dH(y \mid n-1)\right).
\end{aligned}$$

Recall that $\pi_1(y)$ is strictly increasing. This implies that $\pi_1(y) - \int_{\underline{x}}^{\overline{x}} \pi_1(z)\, dH(z \mid n)$ is strictly increasing as well, negative when $y = \underline{x}$, and positive when $y = \overline{x}$. It follows that there is an x_0 such that $\pi_1(y) - \int_{\underline{x}}^{\overline{x}} \pi_1(z)\, dH(z \mid n-1) \leq 0$ for all $y < x_0$, and ≥ 0 for all $y > x_0$. Hence,

$$x \mapsto \int_0^x dH(y \mid n-1) \left[\pi_1(y) - \int_0^{\overline{x}} \pi_1(z)\, dH(z \mid n-1) \right]$$

is first decreasing and then increasing. Since this function is obviously equal to zero at the upper bound of its domain, $x = \overline{x}$, it follows that $H(x \mid n) \leq H(x \mid n-1)$, and we have established first-order stochastic dominance.

Internet Appendix

IA.1 Dealer Excess Inventory Measurement

In this section, we discuss our measure of dealer excess inventory in Equation (4) in more detail. Let $i_b(t)$ denote dealers' inventory of bond b at time t. Moreover, let $e_b(t)$ denote dealers' excess inventory of bond b at time t, which we define as:

$$e_b(t) \equiv \frac{i_b(t) - \mu_b(t)}{\sigma_b(t)},$$

where $\mu_b(t)$ is the desired level of inventory and $\sigma_b(t)$ is the standard deviation of inventory for bond b at time t. To allow for changes in inventory dynamics over time, we estimate excess inventory on a rolling basis. Given an estimation window Δt, an estimator for $\mu_b(t)$ is

$$\hat{\mu}_b(t;\Delta t) = \frac{1}{\Delta t}\int_{t-\Delta t}^{t} i_b(s)ds,$$

and an estimator for $\sigma_b(t)$ is

$$\hat{\sigma}_b(t;\Delta t) = \sqrt{\frac{1}{\Delta t}\int_{t-\Delta t}^{t} \left(i_b\left(s\right) - \hat{\mu}_b\left(t;\Delta t\right)\right)^2 ds}.$$

Unfortunately, the direct implementation of these estimators is challenging since, to the best of our knowledge, data on the *level* of dealer inventory for individual corporate bonds are not available. However, using transaction records on individual corporate bonds from the TRACE database, it is possible to measure *changes* in dealer inventory. With that in mind, let $\Delta i_b(u,s)$ denote the change in inventory between time u and s, $\Delta i_b(u,s) \equiv i_b(u) - i_b(s)$. The enhanced TRACE database provides information on the time of trades, transaction parties (dealer vs. customer), the side of a trade (dealer buys vs. dealer sells), and the traded quantities. Letting S_j denote the trade side of transaction j, $S_j \in$ {Dealer buys, Dealer sells, Interdealer transaction}, letting τ_j denote the time of transaction j, and letting Q_j denote the quantity of the transaction, we calculate the change in dealer inventory between $t < u$ as follows:

$$\Delta i_b(u,t) = \sum_{j|t\le\tau_j\le u} \left(\mathbb{I}_{\{S_j=\text{ Dealer buys}\}} - \mathbb{I}_{\{S_j=\text{ Dealer sells}\}}\right) Q_j.$$

Then, it is straightforward to show that our estimators can equivalently be written as:

$$\hat{\mu}_b(t;\Delta t) = \Delta\hat{\mu}_b(t,\Delta t) + i_b(t-\Delta t),$$
$$\hat{\sigma}_b(t;\Delta t) = \sqrt{\frac{1}{\Delta t}\int_{t-\Delta t}^{t} \left(\Delta i_b(s,t-\Delta t) - \Delta\hat{\mu}_b(t;\Delta t)\right)^2 ds},$$

where $\Delta\hat{\mu}_b(t,\Delta t) = \frac{1}{\Delta t}\int_{t-\Delta t}^{t} \Delta i_b(s,t-\Delta t)ds$. As a result, we obtain two equivalent formulations of the excess inventory estimator:

$$\hat{e}_b(t;\Delta t) = \frac{i_b(t) - \hat{\mu}_b(t;\Delta t)}{\hat{\sigma}_b(t;\Delta t)} = \frac{\Delta i_b(t,t-\Delta t) - \Delta\hat{\mu}_b(t;\Delta t)}{\hat{\sigma}_b(t;\Delta t)}.$$

The last equality does not involve inventory level, only inventory changes. As a result, we implement this version of the estimator using available transaction records at the bond level.

IA.2 Additional Figures and Tables

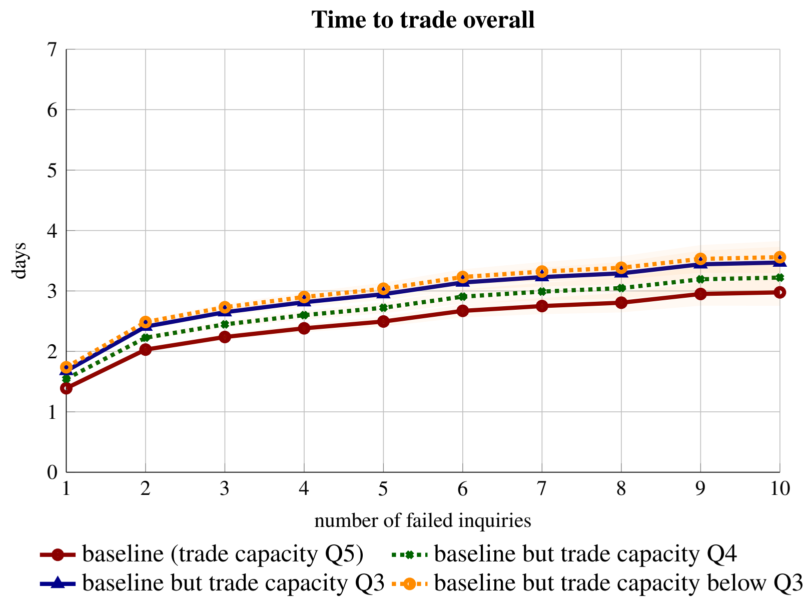


Figure IA.1. Estimated conditional time to trade from the MLE: impact of dealer excess inventories
This figure plots the estimated time to trade from Equation (2), conditional on the number of failed inquiries and dealer trade capacity categories. We rank inquiries based on a measure of "trade capacity": high trade capacity means either high excess dealer inventory for customer purchases or low dealer excess inventory for customer sales. We first sort buy and sell inquiries separately into quintiles based on our measure of dealer excess inventory defined in Equation (4). We then define an indicator for trade capacity quintile 5 (Q5) as (buy Q5 inventory indicator + sell Q1 inventory indicator), and similarly for other quintiles. "Trade capacity Q4" is an indicator for inquiries in the 4$^{\text{th}}$ quintile of the trade capacity measure, and similarly for other "trade capacity" quintiles indicators. The baseline category is an odd-lot purchase of an investment-grade bond, with high turnover, during normal times, for a connected investor, in Q5 of dealer capacity category, after one failed inquiry.

Table IA.1. Trade probabilities: inquiry vs. child order level

This table presents logit regression results of whether trade occurs as the dependent variable and indicators for trade and customer characteristics as independent variables, defined in Tables 5 and 8. Column (1) presents the regression at the inquiry level for trade on MKTX. The corresponding child order level estimates for trade on MKTX or voice are presented in column (2). Heteroskedasticity-robust standard-errors are reported in parentheses.

Dependent Variables:	inq. is traded on MKTX	child is traded on MKTX/voice
Model:	(1)	(2)
Variables		
(Intercept)	2.231***	3.270***
	(0.0032)	(0.0047)
Micro size	0.3344***	0.5321***
	(0.0023)	(0.0031)
Round lot	-0.3437***	-0.5465***
	(0.0040)	(0.0050)
Block trade	-0.7327***	-1.044***
	(0.0117)	(0.0138)
Sell	0.6301***	0.6181***
	(0.0021)	(0.0030)
HY	-0.4462***	-0.1778***
	(0.0028)	(0.0039)
Covid	-0.9289***	-0.7719***
	(0.0063)	(0.0083)
Old age	-0.1935***	-0.2520***
	(0.0023)	(0.0031)
High time-to-maturity	-0.0152***	-0.1402***
	(0.0022)	(0.0030)
Low turnover	-0.2895***	-0.3352***
	(0.0029)	(0.0038)
Low amt outstanding	-0.6594***	-0.7774***
	(0.0022)	(0.0030)
Connected decile < 7	-2.326***	-2.240***
	(0.0029)	(0.0038)
Connected decile 7	-0.6024***	-0.5286***
	(0.0041)	(0.0062)
Connected decile 8	-1.110***	-1.171***
	(0.0036)	(0.0049)
Connected decile 9	-0.2790***	-0.3332***
	(0.0029)	(0.0043)
Fit statistics		
Observations	6,684,638	6,241,870
Squared Correlation	0.17918	0.11526
Pseudo R^2	0.16766	0.16067
BIC	5,772,256.4	3,425,744.2

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table IA.2. Summary statistics
This table presents summary statistics for size, bond age and maturity, rating, and trade direction for all child orders (column 1), all inquiries (column 2), and child orders with at least one failed inquiry (column 3). "Sell" takes the value of 1 for a sale request, and zero otherwise; "HY" takes the value of 1 if the bond is high-yield, and zero otherwise; "Dealer-submitted" takes the value of 1 if the inquiry is submitted by a dealer, and zero otherwise.

	Child orders (all) (1)	Inquiries (all) (2)	Child orders (≥ 1 failed inq.) (3)
HY	0.17	0.18	0.26
Sell	0.52	0.51	0.42
Dealer-submitted	0.10	0.11	0.23
Size			
micro size (< \$100k)	0.49	0.48	0.37
odd lot (\$100k–1 million)	0.42	0.43	0.52
round lot (\$1–5 million)	0.09	0.08	0.10
block trade (> \$5 million)	0.01	0.01	0.01
Bond age distribution			
Average bond age	3.85	3.91	4.43
< 2 years	0.35	0.34	0.31
2–5 years	0.39	0.39	0.38
5–20 years	0.26	0.27	0.30
> 20 years	0.01	0.01	0.02
Bond maturity distribution			
Average maturity	12.43	12.53	13.71
< 2 years	0.002	0.002	0.002
2–5 years	0.07	0.07	0.06
5–20 years	0.73	0.73	0.68
> 20 years	0.20	0.20	0.25
Observations	9,861,143	11,020,815	2,774,478

Table IA.3. Child orders statistics: Inter-arrival times.
This table presents summary statistics about time between child order events (in business days). A child order can be viewed as a sequence of events, as depicted in Figure 1. Each element of the sequence is one of four possible events: an untraded inquiry on MKTX, a MKTX inquiry with trade, a voice trade, and, if the child order ends without a trade, an exit. Columns (1)–(3) present time, in business days, to an untraded inquiry on MKTX, a MKTX trade, and a trade on voice across child orders, conditional on the number of failed inquiries to date.

	Time to MKTX inq. w/o trade (1)	Time to MKTX inq. w trade (2)	Time to voice trade (3)
After 1 failed inquiry	0.82	0.65	1.04
After 2 failed inquiries	0.87	0.82	1.34
After 3 failed inquiries	0.85	0.85	1.46
After 4 failed inquiries	0.84	0.88	1.53
After 5 failed inquiries	0.82	0.85	1.56
After 6 failed inquiries	0.80	0.87	1.56
After 7 failed inquiries	0.78	0.89	1.63
After 8 failed inquiries	0.77	0.84	1.59
After 9 failed inquiries	0.75	0.86	1.44
After 10 failed inquires	0.72	0.88	1.44

Table IA.4. Quote quality for the first inquiry and trade probability
This paper present logit regression estimates. The dependent variable is an indicator for whether the inquiry is traded on MKTX and the main dependent variable is quote quality for the first inquiry in child orders, defined in Equation (3). In columns (1) and (2) we include bond and trade characteristics described in Tables 5 and 8.

Dependent Variable:	inq is traded on MKTX	
Model:	(1)	(2)
Variables		
(Intercept)	0.4834***	
	(0.0128)	
Quote quality	0.2663***	0.3079***
	(0.0006)	(0.0077)
Bond & trade controls	Yes	Yes
Customer controls	Yes	
Fixed-effects		
customer		Yes
issuer		Yes
dealer		Yes
year-month		Yes
Fit statistics		
Observations	6,954,147	7,679,325
Squared Correlation	0.28497	0.43891
Pseudo R^2	0.26354	0.39911
BIC	5,277,697.2	5,290,981.4
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*		