FEDERAL RESERVE BANK
PHILADELPHIA

Working Papers
RESEARCH DEPARTMENT

Sequential Search for Corporate Bonds

Mahyar Kargar
University of Illinois at Urbana–Champaign

Benjamin Lester
Federal Reserve Bank of Philadelphia

Sébastien Plante
University of Wisconsin–Madison

Pierre-Olivier Weill
UCLA, NBER, CEPR, and Visiting Scholar, Federal Reserve Bank of Philadelphia Research Department

WP 25-08

PUBLISHED
March 2025

ISSN: 1962-5361

DOI: https://doi.org/10.21799/frbp.wp.2025.08

Sequential Search for Corporate Bonds*

Mahyar Kargar[†] Benjamin Lester[‡] Sébastien Plante[§]
Pierre-Olivier Weill[¶]

February 20, 2025

Abstract

Customers in over-the-counter (OTC) markets must find a counterparty to trade. Little is known about this process, however, because existing data consist of transaction records, which only reveal the *outcome* of a search. Using data from a trading platform for corporate bonds, we unpack the search process. We analyze how long it takes customers to trade and how dealers' offers evolve across repeated inquiries. We estimate that it takes 2-3 days to complete a transaction after an unsuccessful attempt, with substantial variation across trade and customer characteristics. Our analysis offers insights into the sources of trading delays in OTC markets.

Keywords: Over-the-counter markets, intermediation, corporate bonds, time-to-trade.

JEL Classification: G11, G12, G14, G21.

*We thank Briana Chang (discussant), Tatyana Deryugina, Terry Hendershott, Tim Johnson, Jian Li (discussant), Dmitry Livdan (discussant), Yi Li (discussant), Maureen O'Hara, Christian Opp (discussant), Gábor Pintér (discussant), Kerry Y. Siani (discussant), Alan Shepard, Antoinette Schoar (editor), Kumar Venkataraman (discussant); two anonymous referees and an associate editor; conference participants at AEA/CEANA, AFA, 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.

[†]University of Illinois at Urbana-Champaign. Email: kargar@illinois.edu

[‡]Federal Reserve Bank of Philadelphia. Email: benjamin.lester@phil.frb.org

[§]University of Wisconsin-Madison. Email: splante@wisc.edu

[¶]UCLA, NBER and CEPR. Email: poweill@econ.ucla.edu

Conflict of Interest Disclosure Statement

Mahyar Kargar:
I have nothing to disclose.

Benjamin Lester:
I have nothing to disclose.

Sébastien Plante:
I have nothing to disclose.

Pierre-Olivier Weill:
I have nothing to disclose.

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 consist 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, as far as we know, it is the only available 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 way that dealers respond 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, in Section 2, we document that executing a trade often requires multiple attempts or "inquiries," and explain how we organize the data in a way that captures the sequential nature of investors' search process. Then, in Section 3, we introduce a statistical model and estimate it via maximum likelihood; this exercise yields novel estimates of time-to-trade and describes how these estimates vary across customer, bond, and trade attributes. Lastly, in Section 4, 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 mechanisms that generate trading delays in OTC markets. We now describe each of these steps in greater detail.

To start, we establish that customers often engage in sequential search when attempting to trade. 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 statistical 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. As a benchmark, we consider a fairly liquid trade: an odd-lot purchase of an investment-grade bond with amount outstanding above the median by a customer who is "well connected" to dealers. For this benchmark, we find that it takes approximately two days to complete a trade 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 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 60% 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 unconditional expected time to trade 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 observe how long the customer was searching prior to submitting their first inquiry on MKTX. For example, the customer could have searched through voice channels before the initial MKTX inquiry, or could have been consulting dealer "runs" to determine when it would be worthwhile to make an inquiry.[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 parameters that are typically either fixed to some value deemed reasonable or identified via indirect inference.[2] Second, the correlations we find between 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.

In the third step of our analysis, we dig deeper into the behavior of customers and dealers over

[1]See Hendershott, Li, Livdan, Schürhoff, and Venkataraman (2021) for a more detailed description of the dissemination of dealer runs.

[2]In addition, though we do not analyze order-splitting explicitly, our estimates are nonetheless informative about the trade-off that investors in this market face when deciding whether—and to what extent—they should break up a large order into smaller blocks. In particular, while it's well known that smaller requests can generate better replies from dealers (Kyle, 1985), our estimates of time-to-trade are among the first to quantify the time cost of splitting orders in the corporate bond market.

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 at the press of a button. If one were to take a narrow, literal view of trading frictions, they might anticipate that an electronic trading platform would make the US corporate bond market replicate the outcome of a frictionless exchange. Yet we find that it does not. Instead, we provide evidence that the customers in our sample are facing the trade-off that arises in canonical models of consumer search (McCall, 1970)—namely, the trade-off between trading immediately and continuing to search for a better price.

To start, we document significant dispersion in both the number and the quality of dealers' responses across inquiries, providing customers with incentive to reject "bad" offers and send an identical inquiry hours (or days) later. Then, we show that the distribution of responses does not deteriorate or "move against" the customer over the course of their search, but rather remains fairly stationary. As a result, there are potential gains from search: customers who reject the quotes they receive at the initial inquiry eventually fulfill their trade request at a later inquiry about 40% of the time, and the price they ultimately accept is, on average, about 10 bps better than the best offer they received at the initial inquiry.

But what generates dispersion in the number and quality of dealers' responses? Why do some dealers reply to one inquiry and not the next? Why do they change their quote from one inquiry to the next? We show that fluctuations in dealers' inventory holdings are an important factor in answering these questions. In particular, by linking dealers in MKTX and TRACE, we are able to construct measures of bond-level inventory holdings across dealers. Using these measures, along with information about each dealer's response to customer inquiries, we document that dealers are more likely to reply to a customer's request to purchase (sell) a bond—and more likely to quote a lower ask (higher bid) price—when they are holding a relatively large (small) quantity of the bond in inventory. Hence, given the important role that a dealer's inventory holdings play in their bidding behavior, our results point to *frictions in the interdealer market* as an underlying force in explaining dispersion in the number and quality of quotes available to customers, and thus for the presence of (rational) trading delays in the corporate bond market.

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 quantity of a specific bond; this allows us to document a number of new insights about the sequential nature of customers' search process. Second, 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, and to quantify the gains from search, 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 (2024) 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.[4] Since existing data provides detailed information about dealers' transactions, several studies used the restrictions of search models to identify dealers' time-to-trade. For example, in the models of Afonso and Lagos (2015), Üslü (2019), and Brancaccio and Kang (2021)—in which search is random and the distribution over agents' states is continuous—every meeting results in a trade,

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

[4]More generally, see Pagnotta and Philippon (2018) for a detailed discussion of trading speeds across various markets and trading mechanisms.

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 case, aircraft) for sale at a time. Such information is typically not available in financial 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. 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, it is important to note that our measurement does not impose theoretical restrictions from a specific search-theoretic 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, and the magnitudes we report are vastly different: their findings from the equity market three decades ago are in the order of minutes, whereas our findings from the current corporate bond market are measured in hours and days.

[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.

2 Data

Our main source of data is MarketAxess (MKTX), the leading electronic trading platform in the US corporate bond market. Prior to the introduction of MKTX in 2000, this 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. MKTX offered a trading platform allowing buy-side traders (henceforth customers) to query multiple dealers at once via an electronic request for quote (RFQ), with the objective of reducing the time-consuming process of gathering quotes and 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, the sample of trades on MKTX differs from that of the overall market in several important ways. Panel (a) of Table 1 reports the market share of MKTX across trade size and bond rating categories, while panel (b) provides descriptive statistics of trades that occur on MKTX relative to the total trading volume observed in TRACE. Consistent with earlier findings by O'Hara and Zhou (2021), we find that customers use MKTX relatively more frequently for smaller trades (specifically odd lot), for investment-grade bonds, for bonds with more time until maturity, and to sell (instead of buy).

When requesting a quote on the MKTX platform, customers specify the bond they wish to trade, the desired quantity, the trade direction (buy or sell), and the duration of the auction (usually between 5 and 20 minutes). Once submitted, an inquiry is sent to the customer's list of pre-authorized dealers, i.e., those dealers with which the customer has a pre-existing relationship. On the receiving end, dealers observe the details of the inquiry, including the customer's identity.[7] 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.

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

[7]In 2012, MKTX introduced Open Trading, an all-to-all trading protocol where non-pre-authorized dealers (including other customers) can respond to RFQs. In contrast to the "classic" trading protocol, where the identity of customers and dealers is mutually disclosed, trades under the Open Trading protocol are anonymous: MKTX stands between the inquirer and the responder and conceals the identities of counterparties. However, as Hendershott, Livdan, and Schürhoff (2021) report, the vast majority of trades are still intermediated by pre-authorized dealers who can observe the customer's identity.

[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 match the quantity level requested by the customer.

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 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 trade with a dealer or via other electronic platforms such as Tradeweb or Bloomberg. In what follows, we simply refer to trades occurring outside of MKTX as "voice" trades.[9] As we describe below, we identify these trades using the academic version of the TRACE data set provided by FINRA, which contains detailed reports of every successful trade, whether electronic or voice, along with anonymized dealer identifiers. 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 fundamentals (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 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

[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.

and ask prices at a high frequency (every 15 to 60 seconds).[10] These imputed prices 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 standing CP+ bid and ask prices.[11]

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.[12]

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

[10]The construction of these reference prices follows two steps. First, MKTX trains a machine learning algorithm using three distinct sources of bond trading data: (1) historical TRACE prints; (2) indicative bond price data streamed by dealers; and (3) RFQ responses sent by liquidity providers on the MKTX trading platform. Beyond trading data, MKTX uses bond level information and other broad market data 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.

[11]Relative to other common price benchmarks, such as the last inter-dealer transaction price (used by Hendershott and Madhavan, 2015 and others), CP+ has several advantages. First, CP+ provides a two-sided benchmark (bid and ask), which we exploit in Section 4 when evaluating the quality of dealer quotes. Second, CP+ is available intraday at a high-frequency (every 15 to 60 seconds), whereas other benchmarks are typically updated at significantly lower frequency (up to several days between updates). Lastly, the CP+ algorithm uses additional, non-public data, which can potentially create a more accurate forecast.

[12]In column (9) of 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.

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 the behavior of customers and dealers within each inquiry is potentially informative about the trading process, a careful examination of the data reveals that individual inquiries are often part of larger trading orders. In these instances, we argue that individual inquiries should not 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 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 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 typical feature 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.[13] 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

[13]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.

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.

Second, we construct child orders by grouping 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 a trade, either on MKTX or via voice; (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 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.

As noted above, 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 via voice. These two outcomes have different economic implications: the former indicates that the attempt to trade was not completed (or at least, not under the original terms), while the latter indicates that it was completed but on a different platform. As such, we argue that it is important to distinguish between these two outcomes, which requires matching customer inquiries on MKTX that result in a voice trade using the corresponding TRACE record. Unfortunately, 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 small. 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, quantity, and 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 one closest 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. 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. 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 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. Panel (b) of Table 4 reports that multiple trading attempts are observed in 13% of child orders and account for 14% of trading volume.[14]

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 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 observable 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. Third, we

[14]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 inquiries with multiple attempts to trade. 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.

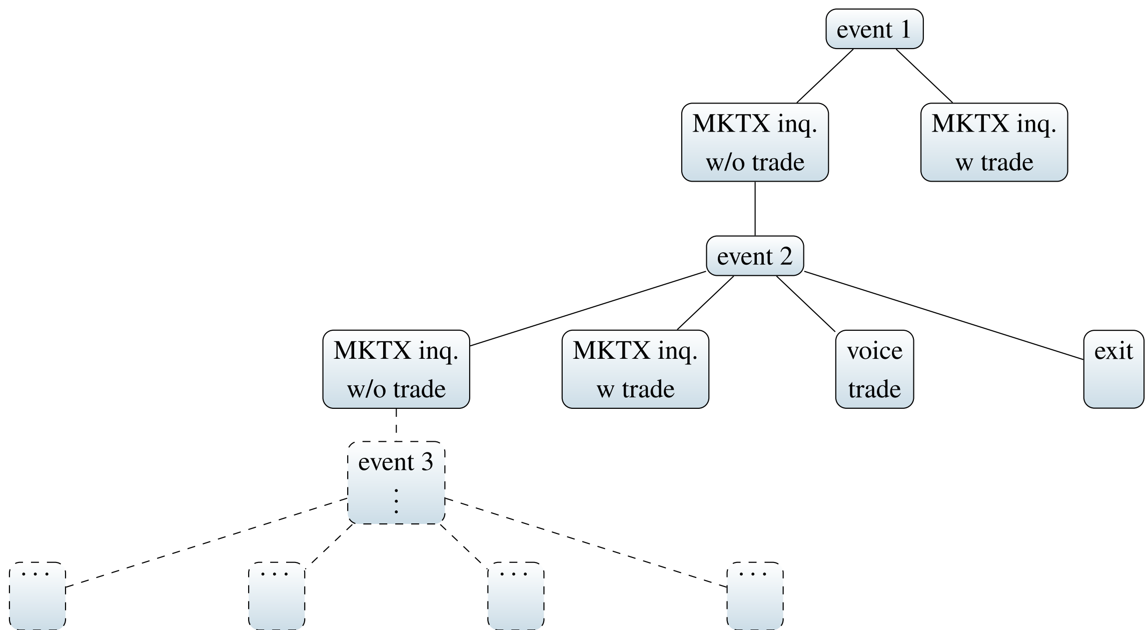

Figure 1. A child order event tree
A child order can be viewed as a sequence of events. Each element of the sequence is one of four possible events: a MKTX inquiry without trade, a MKTX inquiry with trade, a voice trade and, if the child order ends without a trade, an exit. By construction, the first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

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, either by sending an inquiry for a different amount (or different bond) or by abandoning the trade altogether; for the sake of convenience, we call this event an "exit," though it should be understood that it can include continued search for a related trade. 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 orders 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 on MKTX successfully results in trade, whereas column (2) reports the probability that a child order successfully results in trade, 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 period. 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 varies systematically with the characteristics of both the inquiry 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 more likely to be fulfilled relative to larger trades. However, perhaps the most significant source of heterogeneity across inquiries derives from differences across customers. To illustrate this fact, we create a measure of the customer's "connectedness" to proxy for the number of existing relationships with dealers (or other unobserved characteristics that influence the number 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 the 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.[15]

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 (column 1), while in the second regression the dependent variable is whether trade occurred on MKTX or voice at the child order level (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

[15]To check for robustness, we also constructed an alternative measure of customer connectedness based on the number of dealers that a customer *invited* to respond to an inquiry. The main takeaways highlighted above, along with the quantitative estimates of time to trade reported in Section 3, are largely unchanged under this alternative measure of customer connectedness.

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, when the bond being requested had above-median turnover and amount outstanding and below-median time to maturity and age. The results are in Table IA.1 in the Internet Appendix 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).[16] 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.[17] 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 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 from 90% to about 79%, but at the child order level it falls much less, from 96% 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 of child orders significantly differs from its inquiry- or trade-level counterparts in other dimensions as well.[18] 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, size, and bond characteristics. To measure the inter-arrival time of trading

[16]Since the unit of the coefficients is the log odds ratio of trade, the intercept in column (1) of Table IA.1 implies that the probability 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 Figure 2.

[17]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 trades with in a given day.

[18]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.

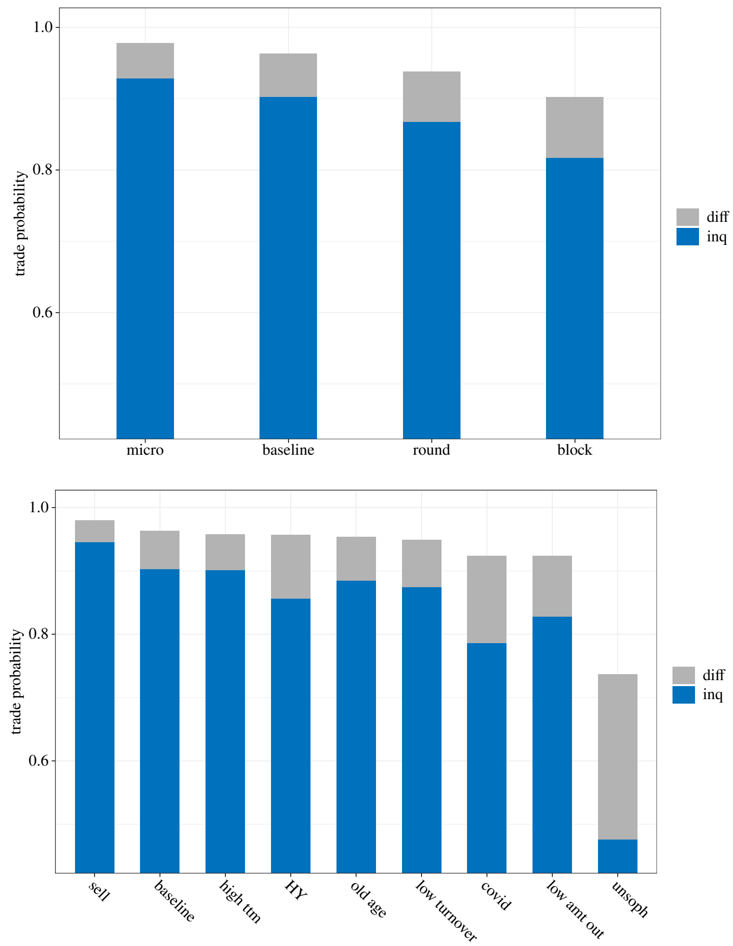

Figure 2. Estimated trade probability on MKTX at inquiry and child order levels
This figure compares the estimated trade probability using logit regression estimates from Table IA.1. The blue bars present trade probabilities (on MKTX) at the inquiry level. The gray bars shows the extra trade probability (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. Inquiries submitted by dealers are excluded.

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 requests to buy, and towards inquiries for high-yield bonds and larger quantities, 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. In Section 3, we argue 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. 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.[19]

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 runs only when the market is open, accounting for the fact that the market is closed at night, on weekends, and during holidays). However, as we argue below, this estimate is clearly biased downwards, since observing this event requires that none of the other events have occurred first.

3 Measuring Time to Trade

In this section, inspired by the event tree in Figure 1, we formulate a statistical model and use maximum likelihood to estimate the time it takes for customers to contact and trade with dealers, either via MKTX or via voice, after a failed initial inquiry. 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.

[19]In the context of search models, "duration dependence" typically refers to the observation that the probability (or arrival rate) of an event depends on how long a subject has been searching.

Using a statistical model to estimate the arrival rates of the different events in the tree of Figure 1 has several advantages. First, it allows us to account for competing risk bias. In particular, observing an event occur—like a trade on MKTX—implies that a *different* event (like a voice trade or an exit) did not occur first. Hence, estimating the average time it would take to trade on MKTX requires adjusting for the fact that the trades we observe in the data were completed sufficiently fast that a different event, or a "competing risk," did not occur first, as otherwise the estimate would be biased downwards. Second, our approach accounts for selection as well: as we establish below, there is substantial heterogeneity across customers, bond attributes, and trade attributes that has first-order effects on the time it takes (and the method that is used) to successfully execute a trade inquiry. Lastly, since our estimates have close counterparts in existing search-based models of OTC markets, our analysis offers guidance for the quantitative values of key parameters and highlights which dimensions of these models fit the data well, and which dimensions must be enriched in order to match certain features of the data.

While the arrival rates of different events are taken as exogenous in this statistical model, in Appendix A we formulate and solve a sequential search model of a child order in the style of McCall (1970). Characterizing optimal behavior in this model generates an endogenous relationship between the arrival rate of meetings between a customer and dealers (via an electronic platform or voice), the distribution of offers the customer receives, and the customer's decision to accept one of these offers or continue searching. This model also allows us to formalize the two sources of bias discussed above—competing risk and selection—and to derive analytical comparative statics. Importantly, the model studied in Appendix A also provides a framework to study selection based on *unobservable* heterogeneity, such as customers' costs from failing to trade, which are not accounted for in the estimates that derive from our reduced-form model.

Finally, it is important to note that our estimates of time to trade are conditional on a failed first inquiry. Calculating the unconditional time to trade would require several adjustments. First, of course, one would have to account for the probability that the trade was successful at the first inquiry; from Table 5, one sees that these probabilities vary across trade and customer characteristics. Second, one would also like to account for search behavior that occurred before submitting an inquiry on MKTX, such as attempting to trade via voice or exploring "dealer runs" containing indicative prices (Hendershott et al., 2021). Unfortunately we have little way of accounting for this second type of adjustment, which is why we focus on time to trade after a failed initial inquiry.

3.1 Statistical model

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 characteristics (such as direction, size, and time), bond characteristics, customer characteristics, *and* the number of failed inquiries on MKTX; the latter is particularly important, in that it allows us to identify potential duration dependence.

Given this framework, conditional on the covariates of the i^{th} event x_i, the event $\omega_i = k$ 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 inquiries, the likelihood function is, evidently:

$$\prod_{i=1}^{n} \left(\sum_k \mathbb{I}_{\{\omega_i = k\}} \lambda(\theta'_k x_i) e^{-\sum_\ell \lambda(\theta'_\ell x_i) \tau_i} \right),$$

where $\mathbb{I}_{\{\omega_i = k\}}$ is an indicator for event k.

Recall that we observe only 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

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

3.2 Maximum likelihood estimation

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 $\hat{\theta}_k = \log(\hat{\pi}_k / \hat{\tau})$.*

Lemma 1 illustrates how to correct for competing risk bias using the arrival rates of other events. In particular, it shows that the observed arrival rate of event k, $\hat{\pi}_k$, has to be adjusted by the inter-arrival times between two events, $\hat{\tau}$.[20]

The estimation results, shown in Table 7, offer guidance about the orders of magnitude of arrival times for different events. For example, the unconditional intensity of a voice trade is $e^{-3.369} = 0.0344$ per business hour, corresponding to an average time of $1/0.0344 = 29.05$ business hours, or about 3.2 business days (assuming 9 hours of trading per day). Importantly, the estimates show that competing risk creates a significant bias in calculating time to trade, as 3.2 business days is much larger than the observed average inter-arrival times shown in Table IA.3 in the Internet Appendix. Similarly, after a failed inquiry, without controlling for trade and customer characteristics, the time to trade (on MKTX or voice) is $1/\left(e^{-4.085} + e^{-3.369}\right) = 19.51$ business hours, or approximately 2.17 business days.

Next, we move to the *conditional* MLE, with controls for trade characteristics. All controls are indicator variables. There is no closed form solution for the estimators. However, since the likelihood function is concave in the vector of coefficients, it can be maximized reliably using standard optimization methods. Estimation results are reported in Table 8.

Table 8 shows how the intensities of each event, $\lambda(\theta_k' x)$, vary with trade characteristics. The

[20] This correction is derived more generally within the context of the sequential search model studied in Appendix A.

intensities for the baseline category are obtained by taking the exponential of the intercept.[21] In particular, the arrival intensity of a baseline trade on MKTX is $e^{-3.37} = 0.0344$ per business hour and the arrival intensity of a voice trade is $e^{-3.18} = 0.0416$ per business hour, for an overall trade intensity of 0.0760 per business hour (approximately once every 1.5 day) on either venue.[22]

The marginal effect of trade characteristics on trade intensity is given by the exponential of the corresponding coefficient. For instance, the coefficients in the fifth row of Table 8 imply that, during COVID, trading intensity was lower on MKTX ($e^{-3.37-0.573} = 0.0194$) and over voice ($e^{-3.18-0.264} = 0.0319$); taken together, this implies a 32% reduction in overall trading intensity relative to normal times. This finding provides evidence 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 effect of sudden selling pressure on market quality.

Table 8 also shows that trading intensity is decreasing in trade size. In particular, the estimated trading intensity (on either venue) of micro trades (less than $100,000) is 26% higher than for baseline odd lot size trades (between $100,000 and $1 million). Moreover, the trading intensity for round lot trades (between $1 million and $5 million) and block trades (larger than $5 million) are 20% and 31% lower, respectively, than they are for odd lots.

We also find that trade intensity depends on the side of a transaction, with the intensity of customer sales being 52% higher than purchases. 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 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.

Rows 13 to 16 of Table 8 show that our measure of customer connectedness is associated with significant heterogeneity in trading intensity on MKTX. Relative to the most connected customers (in the top decile), the trading intensity of customers in the bottom 6 deciles is 39% lower. However, the relationship between trade intensity and connectedness is non-monotonic: our estimates suggest

[21] As a reminder, the (omitted) baseline category is composed of buy requests for an odd lot quantity, submitted by a customer in the top decile of our connectedness measure, during normal times, for an investment grade bond with above-median turnover and amount outstanding and below-median time to maturity and age.

[22] Note that these trading intensities cannot readily be converted to time to trade in the same way we converted the unconditional MLE estimates above. We explain why in Section 3.3 and provide a recursive formula to convert our conditional trading intensity estimates into time to trade.

that customers in the seventh decile trade at a higher intensity than customers in the eighth decile. We expand further on the possible causes of this non-monotonicity in Section 3.3.

Bond characteristics also influence trading intensity. In particular, we find that large and recently issued bonds, along with bonds that have high turnover, are traded at a higher intensity; these findings corroborate the conventional wisdom that bonds with such characteristics are more liquid. Our estimation also reveals that high yield bonds are traded at a lower intensity on MKTX, compared to investment grade bonds. Interestingly, we find that high yield bonds trade with higher intensity on voice, relative to investment grade bonds. However, it is important to remember that our sample includes only trades that originate with an inquiry on MKTX. Hence, this finding could indicate that investment grade bonds that fail to trade on MKTX are particularly hard to trade, while the same is not true of high yield bonds.

The last rows of Table 8 show that, after controlling for trade characteristics, the number of failed inquiries retains predictive power for the intensity of each event. Note that the intensity of an inquiry on MKTX that does not 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 studied in Appendix A, this evidence suggests a role for unobserved heterogeneity at the child order level. 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.

3.3 Time to trade estimation

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. Whereas calculating time to trade using estimates from the unconditional MLE was relatively simple, the dependence of intensities on the number of failed inquiries in the conditional MLE requires a slightly more complicated calculation.

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:

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

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 term 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

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

We can use this formula to calculate the time to trade. Moreover, differentiating (1) 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. Selling a bond takes about half as much time as purchasing a bond, whereas time-to-trade nearly doubled in March 2020 at the outset of the Covid-19 crisis.

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).

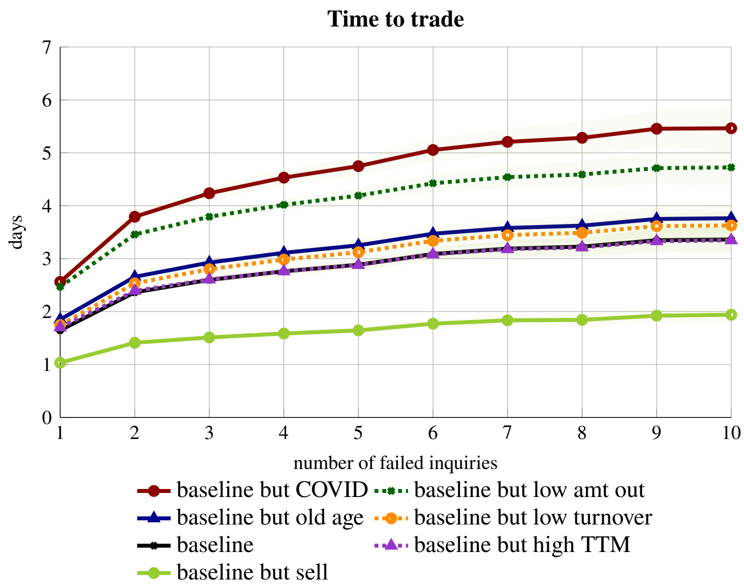

Figure 3. Estimated conditional time to trade after $n \geq 1$ failed inquiries
This figure plots the estimated time to trade from equation (1), conditional on the number of failed inquiries and on several observed trade characteristics, holding the trade size and customer connectedness categories fixed at the baseline levels. "Sell" takes the value of 1 for a sale request, and zero otherwise; "COVID" takes the value of 1 if the inquiry 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 composed of buy requests for an odd lot quantity, submitted by a customer in the top decile of our connectedness measure, during normal times (i.e., not COVID), for an investment grade bond with above-median turnover and amount outstanding and below-median time to maturity and age. Trade size categories are defined in Table 1.

Figure 5 shows that less connected customers, classified as customers who receive fewer offers from dealers, trade much slower. For example, in the baseline category, the most connected customers (in the tenth decile of our measure of connectedness) trade after approximately 2.2 days following two failed inquiries. For the least connected customers, in deciles 1 to 6, it takes almost 3 times longer to trade.[23]

[23] Our estimates also indicate that the effects are non-monotonic, which can be consistent with theory. In particular, in the sequential search model studied in Appendix A, we derive the reservation price such that a customer buyer (seller) is willing to accept an offer only if the price is below (above) their reservation price. Holding a customer buyer's reservation price constant, an increase in connectivity implies that she will receive more (or better) offers, so she is more likely to obtain one that falls below her reservation price and trade. However, since she expects to receive more offers in the future, her reservation price falls and reduces the probability of trading.

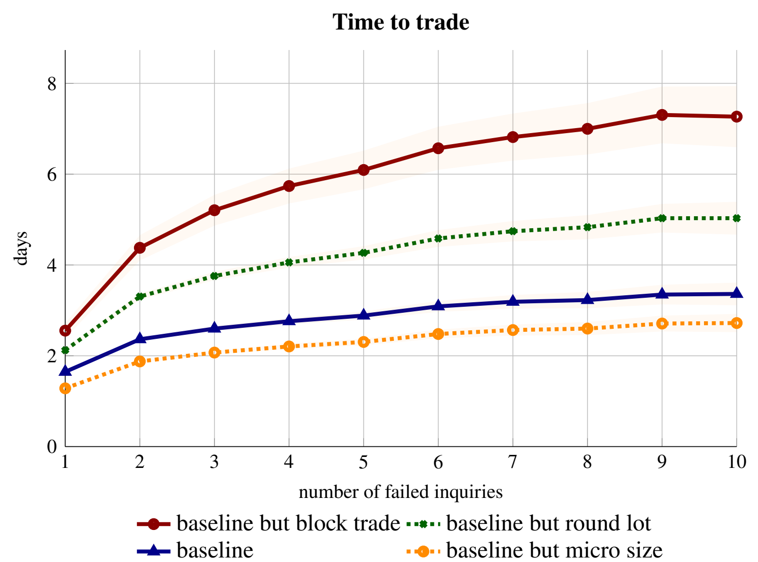

Figure 4. Estimated conditional time to trade: the impact of trade size
This figure plots the estimated time to trade from equation (1), conditional on the number of failed inquiries and trade size categories, holding other trade and customer characteristics fixed at baseline levels. Trade size categories are defined in Table 1. The baseline category is defined in Figure 3.

In Figure 6, we compare time to trade on MKTX and 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.

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

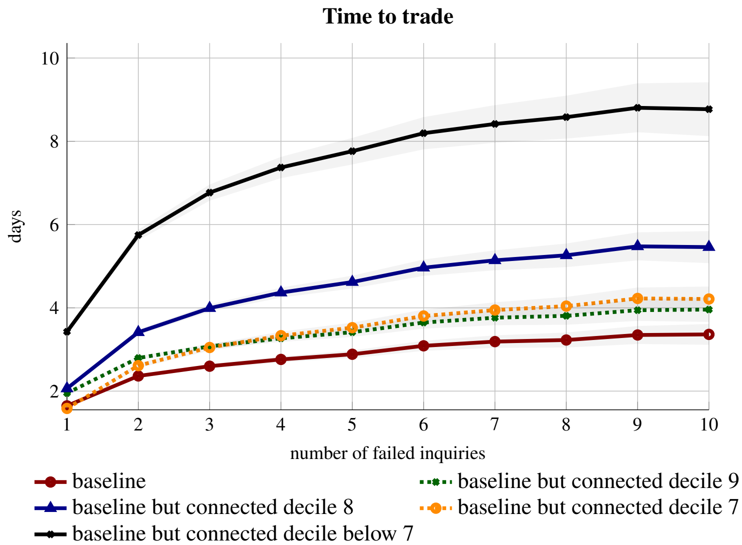

Figure 5. Estimated conditional time to trade: the impact of customer connectedness
This figure plots the estimated time to trade from equation (1), conditional on the number of failed inquiries and on customer connectedness, holding other trade characteristics fixed at baseline levels (defined in Figure 3).

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 phone or push a button. Indeed, we find that a customer usually *can* trade soon after sending an inquiry: in 90% of failed inquiries, the customer received at least one reply but chose not to trade. However, as we document below, rejecting the available offers and searching again for a better quote often yields better prices.

In particular, we first establish that customers face conditions quite similar to those in McCall's (1970) classic model of search: there is significant dispersion in the quality of dealers' responses across inquiries, and the distribution of offers remains fairly stationary as customers search. As a result, customers face the usual trade-off between trading immediately and waiting to find a better price. Those customers who choose not to trade at the initial inquiry ultimately trade at a later inquiry about 40% of the time, achieving a level of price improvement of approximately 10 bps.

Then, we dig deeper to understand *why* there is dispersion in quotes across inquiries. To start, we show that the number of dealers who reply to an inquiry is stochastic, so that dealers

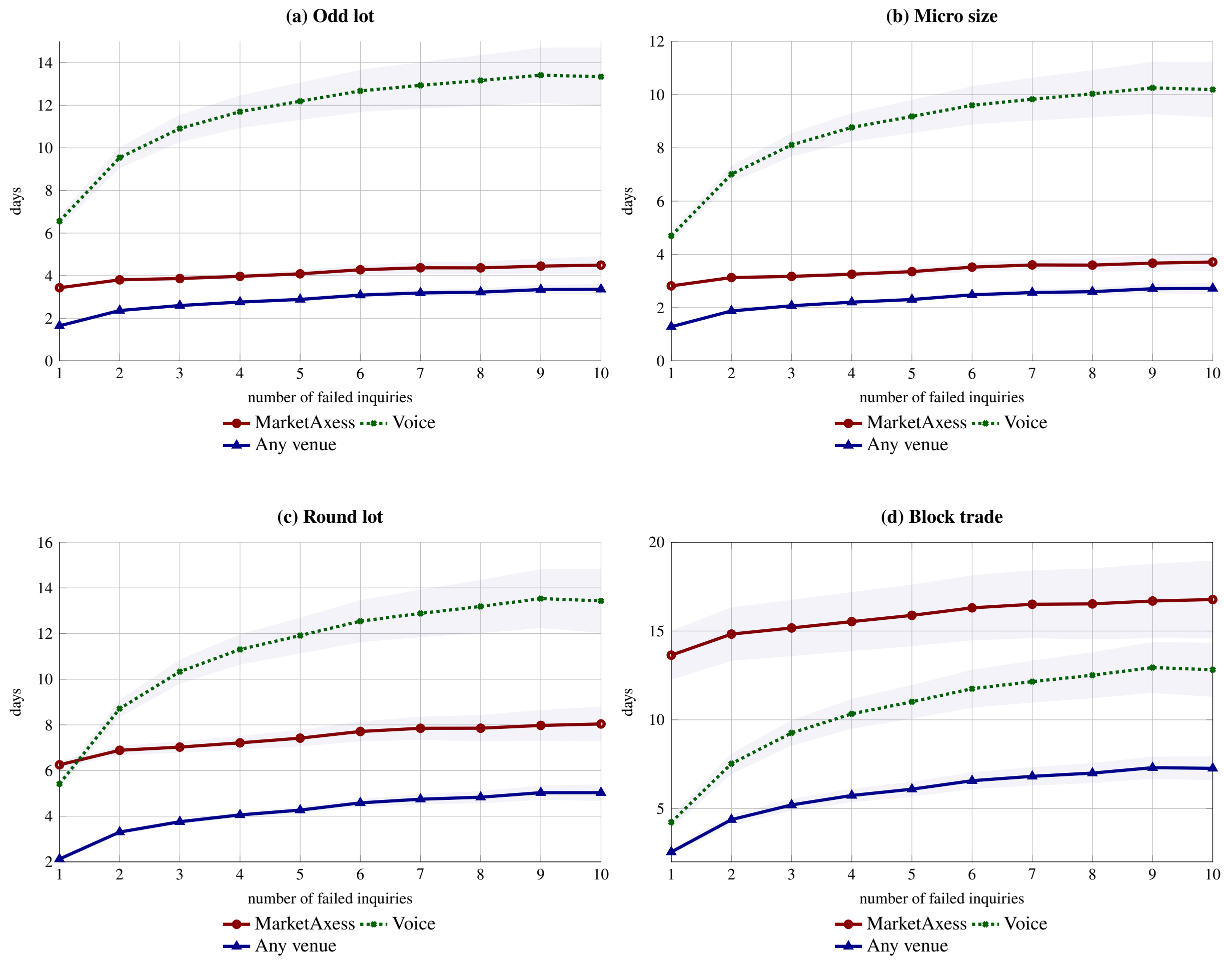

Figure 6. Estimated conditional time to trade: MarketAxess vs. voice
This figure compares the estimated time to trade from equation (1) conditional on the number of failed inquiries in MKTX vs. voice for different size categories holding the customer connectedness and other trade characteristics fixed at the baseline levels. The baseline category is defined in Figure 3 and trade size categories are defined in Table 1.

are uncertain about the amount of competition they face when formulating a quote; as Burdett and Judd (1983) established, this can generate price dispersion even without heterogeneity across dealers. However, we also show that there is an important dimension of heterogeneity that affects both dealers' willingness to respond to an inquiry and the quality of their response—namely, their current inventory holdings of the bond being requested.[24] This finding suggests that a potentially

[24]See Friewald and Nagler (2024), who study the relationship between dealers' inventory and prices using only transactions observed in TRACE.

important source of trading delays in OTC markets are frictions within the inter-dealer sector that prevent dealers from quickly reallocating bonds among each other, which has been studied recently in Hugonnier, Lester, and Weill (2020) and Eisfeldt, Herskovic, and Liu (2024).

4.1 Quote quality, dispersion, and the probability of trade

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 the quotes that a customer receives in response to an inquiry, and study the relationship between this measure of quote quality and the probability of trade.

Consider a child order i with $N \geq 1$ inquiries, and let $p_{i,j,d}$ denote the quote of dealer $d \in \mathcal{D}_{i,j}$ in child order i at inquiry $j \in \{1, ..., N\}$, where $\mathcal{D}_{i,j}$ represents the set of dealers that responded to inquiry j. We define the markup/markdown offered by dealer d as

$$M_{i,j,d} = \begin{cases} \log\left(\frac{p_{i,j,d}}{a_{i,1}}\right) & \text{if customer buy,} \\ \\ -\log\left(\frac{p_{i,j,d}}{b_{i,1}}\right) & \text{if customer sell,} \end{cases} \tag{2}$$

where $a_{i,1}$ and $b_{i,1}$ are the CP+ ask and bid prices, respectively, at the time of the first inquiry in child order i. Hence, the best or "highest quality" quote that the customer receives at inquiry j of child order i can be defined as the smallest markup (or markdown) they are offered:

$$M_{i,j} = \min_{d \in \mathcal{D}_{i,j}} M_{i,j,d}. \tag{3}$$

Note that, for this measure to be well defined, we restrict attention to the sample of inquiries with at least one dealer response, i.e., with $\mathcal{D}_{i,j} \neq \emptyset$.[25]

Table 9 reports the distribution of markups across child orders for the first inquiry—that is, the distribution of $M_{i,1}$ across i—and the relationship between these markups and the probability that

[25]In what follows, we will simply refer to $M_{i,j}$ as the "markup," though it should be understood that it is, indeed, a markdown for customer-sell requests.

the customer accepts. We highlight two key takeaways. First, as expected, customers are likely to accept when the best quote is a good one, and more likely to reject (and, often, continue searching) when it is not. Specifically, we find that customers trade about 85% of the time when the markup offered in their first inquiry is in the first few deciles of the distribution, but this probability falls by nearly 30 percentage points when the markup is in the last decile.[26] Second, Table 9 reveals that there is substantial dispersion in the quality of quotes that customers receive. This finding suggests that there could be gains from continuing to search if the first inquiry does not yield a good quote. In the next section, we attempt to quantify the potential gains (or losses) that accrue to customers from searching.

4.2 Price improvement

To quantify the potential returns to search, we study how the markup changes across inquiries within a child order. Specifically, we quantify the change in the best available price between inquiry 1 and inquiry j, taking into account heterogeneity that arises because of observable bond, customer, or trade characteristics, by running a regression of the form:

$$M_{i,j} = \sum_{j'=2}^{N} \beta_{j'} \, \mathbb{I}_{\{j'=j\}} + \text{controls}_{i,j} + \varepsilon_{i,j}, \tag{4}$$

for various specifications of controls (or fixed effects) at the child order and/or inquiry level, where $\mathbb{I}_{\{j'=j\}}$ is an indicator function that takes on the value of 1 when $j' = j$. Importantly, since inquiry 1 is omitted, the coefficient β_j captures the average change in markup/down between inquiry 1 and inquiry j. Note from equation (2) that, since we are normalizing the markup of each inquiry by the CP+ price at the first inquiry, the change in the markup between inquiry j and inquiry 1 is also equal to the change in the best available price.

Column (1) of Table 10 reports the estimated values of β_j, controlling for certain trade characteristics (such as size and direction) along with customer, bond, and time fixed effects. The results suggest a positive relationship between the level of markups and the number of inquiries in a child order, which could be interpreted as a form of duration dependence: with each additional

[26]Of course, markups 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.

inquiry by a customer, dealers could become less likely to reply to a customer or the quality of their offer could deteriorate, as in the "ringing-phone curse" described in Zhu (2011).

However, these results could also be the byproduct of unobserved heterogeneity at the child order level. For example, suppose some child orders occur at a time when dealers are less willing to buy or sell that particular bond, for reasons that are not explained by observable characteristics of the bond, the trade request, the customer, or time fixed effects. If child orders that generate many/better replies are more likely to result in trade, then the composition of child orders will skew towards those that generate fewer/worse replies as j increases, and markups will appear to be increasing in the number of inquiries.

To account for unobserved heterogeneity across child orders, column (2) repeats the same regression analysis with child order fixed effects rather than customer, bond, and time fixed effects. Under this specification, the best offer at inquiry $j \in \{2, 3, 4\}$ is actually slightly *better* than at the first inquiry—by about 2-3 bps—and otherwise little changed. Hence, we do *not* find strong evidence of duration dependence *within* a child order: conditional on rejecting $p_{i,1}$, the average price at future inquiries within child order i is fairly constant.

Since the distribution of quotes does not appear to move against the customer over the course of a child order, but rather to remain fairly stable (as in the classic model of McCall, 1970), there are potential gains from search provided that there is sufficient dispersion in the quality of offers. To study the extent to which customers find better quotes after rejecting the offers at the first inquiry, we consider a regression of the form

$$M_{i,j} = \sum_{j'=2}^{N} \beta_{j'} \mathbb{I}_{\{j'=j\}} + \sum_{j'=2}^{N} \gamma_{j'} \left[\mathbb{I}_{\{j'=j\}} \times \mathbb{I}_{\{\text{inquiry } j' \text{ is traded}\}} \right] + \text{controls}_{i,j} + \varepsilon_{i,j}, \tag{5}$$

where $\mathbb{I}_{\{\text{inquiry } j' \text{ is traded}\}}$ is an indicator function that takes on the value 1 if the customer traded at inquiry j' on MKTX. Columns (3) and (4) report the estimated values of the coefficients β_j and γ_j, which summarize the average change in markup between inquiry 1 and inquiry j for child orders that are successful and those that are unsuccessful.[27] Focusing on column (4), which controls for child order fixed effects, we find that customers who trade at the second inquiry experience, on

[27]Note that, by studying price changes within child orders that are unsuccessful, this specification accounts for the opportunity cost of non-execution—an important component of trading costs that has been studied extensively in equity markets using the implementation shortfall approach (see, e.g., Perold, 1988; Conrad, Johnson, and Wahal, 2001).

average, a price improvement of $\beta_2 + \gamma_2 \approx 11$ bps relative to the best offer at the first inquiry.[28] We find similar levels of price improvement for trades that occur after more than two inquiries, with a slight decline for inquiries $j \geq 8$.

Of course, rejecting the best offer at the first inquiry doesn't always pay off: among those who do not trade, the average price at inquiry $j \geq 2$ is worse than the offer they received at the first inquiry. Hence, our results elucidate the tradeoff that customers face when deciding whether or not to accept an offer, since waiting for a better offer *takes time* and *carries risk*, in the sense that the customer is not assured of receiving a better offer.

4.3 The origins of quote quality dispersion

We established above that there is considerable heterogeneity in the quality of quotes a customer receives, and that continuing to search can generate better offers. But 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.[29]

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 7, there is considerable dispersion. Of particular importance—from the point of view of the Burdett and Judd (1983) model— is that, in 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.[30]

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 8, 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,

[28]This price improvement is economically significant compared to one-way execution costs from TRACE data for typical institutional-sized corporate bond trades, which range from 20 to 62 basis points (see Table III in Bessembinder et al., 2018).

[29]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.

[30]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).

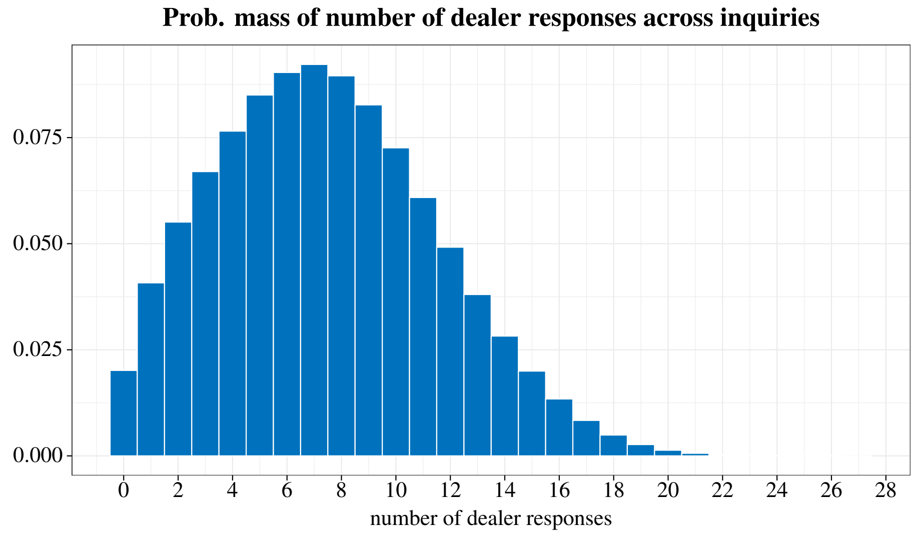

Figure 7. Probability mass function for the number of dealer responses
This figure shows the probability mass for the total number of dealer responses (disclosed and anonymous) at the inquiry level. The sample includes all inquiries submitted by customers.

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.[31]

Finally, linking the two sets of results above, one may wonder whether price 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 price 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.

4.4 Dealers' offer strategy and inventory holdings

Our results suggest that dealers respond to inquiries probabilistically and, when replying to multiple inquiries within a child order, often change the quality of their offer across inquiries. This raises two questions. First, why do dealers selectively respond to some inquiries within a child order but

[31]One may wonder whether the number of responses to an inquiry varies over the course of a child order—that is, whether the *extensive margin* of dealers' replies depends on the number of customer inquiries. Consistent with the results on price improvement, we do not find an economically significant change in the number of dealer replies once we control for child order fixed effects; see Table IA.5 in the Internet Appendix

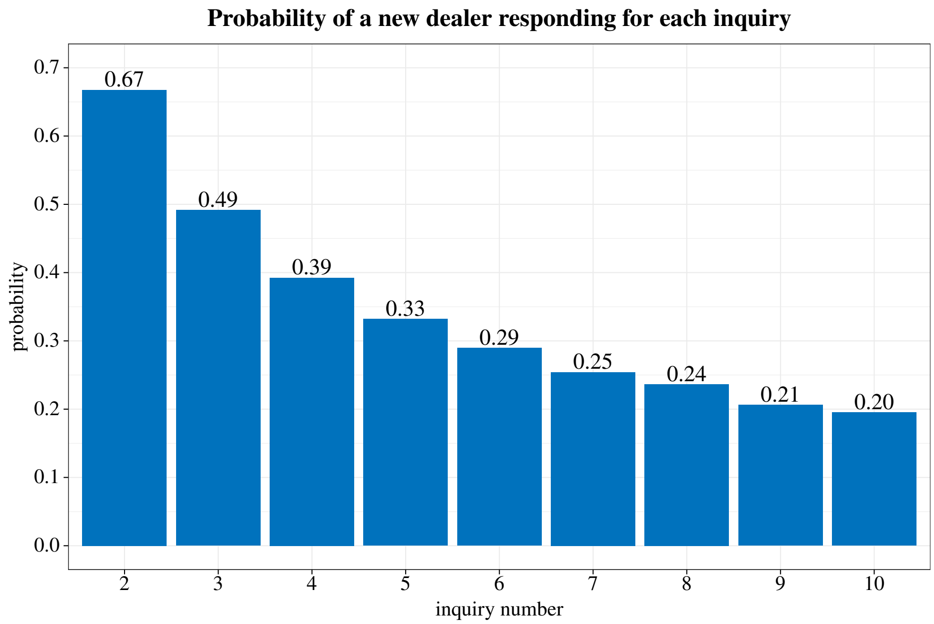

Figure 8. Probability of a response from a new dealer
This figure shows the probability that a new dealer responds to a customer inquiry for a given inquiry number in child orders. A new dealer for inquiry j is defined as a dealer responding to that inquiry for the first time during the entire child order. The sample includes all customer inquiries that receive at least one response from dealers.

not others? Second, what explains variations in a dealer's quotes across inquiries of a child order? A natural explanation is that a dealer's capacity to buy or sell a bond, and willingness to offer a good price, depends on her current inventory holdings.[32]

To test this hypothesis, we construct a direct measure of dealers' excess holdings of each bond in our sample, and use it to study the effect of inventory on the probability that a dealer responds to a customer inquiry and the competitiveness of the quote (conditional on responding).[33] Our analysis allows for inventory to potentially affect the probability and quality of a dealer's response in two ways. First, we study the dealer's excess holdings of individual bonds, which captures the dealer's exposure to the inquiry-specific security. Second, we study the dealer's excess holdings of all bonds, which captures the overall state of the dealer's balance sheet.

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

[33]Since our data do not explicitly provide the list of dealers who received a customer's inquiry, we must infer it: the procedure by which we do so is described in detail in Appendix IA.1.2. In broad terms, we assume that a dealer received a customer's request to trade a given bond if he (1) responded to at least one inquiry from the customer in the previous 120 days, and (2) has traded the bond at least five times (on MKTX or voice) in the last 120 days.

Following Hansch, Naik, and Viswanathan (1998) and Friewald and Nagler (2024), we define the excess inventory holding of dealer d at time t as

$$EI^{\ell}_{d,t} = \frac{I^{\ell}_{d,t} - \text{moving avg. of } I^{\ell}_{d,t} \text{ over the last 120 days}}{\text{moving std. dev. of } I^{\ell}_{d,t} \text{ over the last 120 days}}, \tag{6}$$

where $I^{\ell}_{d,t}$ is the level of dealer d's inventory at time t at the bond-level (ℓ = bond) or across all bonds (ℓ = aggregate). Intuitively, at either the bond-level or at the aggregate balance sheet-level, this measure captures the dealer's current holdings relative to their "typical" inventory holdings. One may wonder how it is possible to calculate this excess inventory measure based on transaction data: indeed, adding up trades allows us to calculate only *changes* in inventories between some initial date and the current date ($I^{\ell}_{d,t} - I^{\ell}_{d,0}$) but not the level ($I^{\ell}_{d,t}$). However, one can easily see that our excess inventory measure (6) is invariant to the initial condition $I^{\ell}_{d,0}$. Hence, it can be equivalently calculated based on the change in inventory, $I^{\ell}_{d,t} - I_{d,0}$, instead of the level, $I^{\ell}_{d,t}$.

We calculate the change in dealers' inventory holdings, $I^{\ell}_{d,t} - I^{\ell}_{d,0}$, as follows (see Appendix IA.1 for details). First, we map the anonymized dealer identifiers in the Academic version of the TRACE corporate bond data to those in MKTX by matching realized transactions on MKTX to their corresponding TRACE record. Using this mapping, we can track *all* of a dealer's transactions, i.e., those executed on the MKTX platform and those conducted through other channels. From those transaction records, we can calculate changes in a dealer's inventory holdings, $I^{\ell}_{d,t} - I^{\ell}_{d,0}$, throughout our sample period by adding dealer purchases and subtracting dealer sales from time zero to time t. We obtain bond-specific change in dealer inventory holdings by adding and subtracting the dealer's purchases and sales in a specific bond, and we calculate measures of aggregate cumulative inventory changes by summing inventory changes of a dealer across *all* CUSIPs.

Using measure (6), we test the effect of a dealer's excess inventory holdings on their probability of responding to an inquiry (the extensive margin) and on the quality of their response (the intensive margin). Specifically, to study the extensive margin, we run the following regression:

$$\begin{aligned}\text{Dealer Responded}_{i,j,d} = {} & \beta_1\, EI^{\text{bond}}_{i,j,d} + \beta_2\, EI^{\text{bond}}_{i,j,d} \times \text{Sell} + \beta_3\, EI^{\text{aggregate}}_{i,j,d} + \beta_4\, EI^{\text{aggregate}}_{i,j,d} \times \text{Sell} \\ & + \text{controls}_{i,j,d} + \epsilon_{i,j,d},\end{aligned} \tag{7}$$

where *Dealer Responded*$_{i,j,d}$ is a binary indicator that equals 1 if dealer d responded to inquiry j of child order i, and zero otherwise, while *Sell* is a binary indicator that equals one in the case of a customer-sell inquiry. The variables $EI^{\text{bond}}_{i,j,d}$ and $EI^{\text{aggregate}}_{i,j,d}$ represent our measures of dealer $\times$ bond (CUSIP) and dealer aggregate excess inventories, respectively, for dealer d's response to inquiry j of child order i. We also include controls for various trade characteristics as before. The hypothesis is that a dealer with higher levels of excess inventory would be more likely to respond to a customer-buy inquiry—since she should have incentive to unload some of her holdings—and less likely to respond to a customer-sell inquiry—since she should be less willing to accumulate even higher levels of excess inventory. If this hypothesis is true, we should find that β_1 and β_3 are positive, and $\beta_1 + \beta_2$ and $\beta_3 + \beta_4$ are negative.

Table 11 presents the extensive margin inventory results. Columns (1) and (2) report estimates from ordinary least square (OLS) regressions with different levels of fixed effects. The coefficient estimate of 0.0326 for β_1 in column (2) with child order fixed effects means that, for a customer-buy inquiry, a one standard deviation increase in a dealer's excess inventory of that particular bond increases the probability of receiving a response from that dealer by 3.26 percentage points. To put this number in perspective, when we restrict our samples to dealers with at least 5 cumulative responses over the 120-day period from the inquiry date, dealers' unconditional propensity to respond is approximately 39%. This implies that a one standard deviation increase in a dealer's excess inventory per CUSIP increases the odds of receiving a response from that dealer for a purchase inquiry by approximately 15%.[34] For a customer-sell inquiry, the same increase in a dealer's excess inventory holdings of that specific bond reduces the probability that the dealer responds by $\beta_1 + \beta_2 = 0.0331 - 0.0461$, or approximately 1.3 percentage points.[35]

Columns (3) and (4) of Table 11 report the estimates of the same coefficients from a logit model with different degrees of fixed effects. To interpret these results, note that the coefficient of 0.1696 for β_1 in column (4) implies that one standard deviation increase in a dealer's excess inventory of that specific bond is associated with a $e^{0.1696} - 1 \approx 18\%$ increase in the odds of responding to the purchase inquiry, which is similar in magnitude to the OLS estimate in column (2). Alternatively, for a customer-sell order, a one standard deviation increase in a dealer's excess inventory of that

[34]To understand how this number is calculated, consider that if we assume the probability of a dealer responding to an inquiry, denoted as p, is at the unconditional average of 0.39, then an increase of 0.0326 in this probability implies that the odds ratio of responding, defined as $p/(1-p)$, increases from $.39/.61 = 0.64$ to $.42/.58 = 0.73$, representing an approximate 15% increase.

[35]This implies that the odds ratio of a dealer responding to a customer-sell inquiry declines by approximately 5% for a one standard deviation increase in the excess inventory of that particular bond.

specific bond reduces the odds of responding by $1 - e^{(0.1696-0.2339)} \approx 6.2\%$, which is again similar to the corresponding results from the OLS regression.

The effects of a dealer's *aggregate* excess inventory holdings on their probability of responding to an inquiry are more muted. In particular, the regression results in columns (1) and (3) indicate no significant relationship between a dealer's overall balance sheet position and their likelihood of responding to either a customer-buy or customer-sell order, while the results in columns (2) and (4)—which allow for child-order fixed effects—report a modest, yet positive relationship.

To study the intensive margin, we regress the spread that a dealer offers in response to an inquiry on the dealer's excess inventory holdings. More precisely, we run the following regression:

$$\begin{aligned} M_{i,j,d} = {} & b_1\, EI^{\text{bond}}_{i,j,d} + b_2\, EI^{\text{bond}}_{i,j,d} \times \text{Sell} + b_3\, EI^{\text{aggregate}}_{i,j,d} + b_4\, EI^{\text{aggregate}}_{i,j,d} \times \text{Sell} \\ & + \text{controls}_{i,j,d} + \varepsilon_{i,j,d}, \end{aligned} \tag{8}$$

where $M_{i,j,d}$ is the markup offered by dealer d in response to inquiry j of child order i, defined in equation (2). Again, the hypothesis is that a dealer with excess inventory will offer a better (worse) quote to a customer-buy (customer-sell) inquiry, in the form of a lower (higher) spread. Under this hypothesis, b_1 and b_3 would be negative, and $b_1 + b_2$ and $b_3 + b_4$ would be positive.

Table 12 reports the intensive margin inventory results. Dealers make better offers for customer-buy (customer-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 in column (2) of Table 12 for b_1 and b_2 imply that a one standard deviation increase in a dealer's excess inventory of the inquiry-specific bond is associated with a 1.5 bps decrease in the markup to a customer-buy inquiry and approximately a 1.1 bps increase in the total markup to a customer-sell inquiry. As in our extensive margin analysis, the effects of a dealer's aggregate excess inventory holdings are less significant.

To summarize, consumers appear to engage in time-consuming searches because they face a nondegenerate, but relatively stationary distribution of replies to successive inquiries within a child order. The stochastic nature of dealers' replies could depend on many factors, but we find that one important consideration is inventory: dealers who currently hold relatively high levels of excess inventory are more willing to sell and less willing to buy, whereas dealers who currently hold relatively low levels of excess inventory are more willing to sell and less willing to buy. Taken together, our results imply that frictions within the dealer sector that prevent the immediate reallocation of excess inventory from one intermediary to another could be a key underlying friction generating the preconditions for search and, in equilibrium, the trading delays that we observe in OTC markets.[36]

5 Conclusion

In this paper, we use data from a leading electronic trading platform to provide new and direct empirical evidence about time to trade 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 often find better prices. 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. 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.

[36]In the presence of a frictionless inter-dealer market, the marginal value of inventories would be constant across dealers and equal to the inter-dealer market price, even if dealers' idiosyncratic inventory holdings fluctuated. As a result, dealers' response behavior would be independent of their idiosyncratic inventory fluctuations, which is inconsistent with our empirical evidence. Generating the observed dependence between dealers' response behavior and their current inventory holdings requires some type of friction in the inter-dealer market, preventing the equalization of dealers' marginal value of inventories. Examples include search models of inter-dealer markets such as Afonso and Lagos (2015) and Hugonnier, Lester, and Weill (2020), and network models such as Malamud and Rostek (2017), Eisfeldt, Herskovic, Rajan, and Siriwardane (2023), and Colliard, Foucault, and Hoffmann (2021). Going forward, an important direction for future research is to identify the key frictions that prevent the equalization of dealers' marginal value of inventories.

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, Mahyar Kargar, Benjamin Lester, and Pierre-Olivier Weill, 2023, Inventory, market making, and liquidity in OTC markets, *Journal of Economic Theory* forthcoming.

Colliard, Jean-Edouard, Thierry Foucault, and Peter Hoffmann, 2021, Inventory management, dealers' connections, and prices in over-the-counter markets, *Journal of Finance* 76, 2199–2247.

Conrad, Jennifer S., Kevin M. Johnson, and Sunil Wahal, 2001, Institutional trading and soft dollars, *Journal of Finance* 56, 397–416.

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.

Eisfeldt, Andrea L., Bernard Herskovic, and Shuo Liu, 2024, Interdealer price dispersion and intermediary capacity, Working Paper 32998, National Bureau of Economic Research.

Eisfeldt, Andrea L, Bernard Herskovic, Sriram Rajan, and Emil Siriwardane, 2023, OTC intermediaries, *The Review of Financial Studies* 36, 615–677.

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

Friewald, Nils, and Florian Nagler, 2024, Dealer inventory and the cross-section of corporate bond returns, *Economics Letters* 239, 117710.

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.

Glebkin, Sergei, Bart Zhou Yueshen, and Ji Shen, 2023, Simultaneous multilateral search, *Review of Financial Studies* 36, 571–614.

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.

Hansch, Oliver, Narayan Y. Naik, and S. Viswanathan, 1998, Do inventories matter in dealership markets? Evidence from the London Stock Exchange, *Journal of Finance* 53, 1623–1656.

Hendershott, Terrence, Dan Li, Dmitry Livdan, and Norman Schürhoff, 2024, When failure is an option: Fragile liquidity in over-the-counter markets, *Journal of Financial Economics* 157, 103859.

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, *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.

Malamud, Semyon, and Marzena Rostek, 2017, Decentralized exchange, *American Economic Review* 107, 3320–3362.

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.

Perold, Andre F., 1988, The implementation shortfall: Paper versus reality, *Journal of Portfolio Management* 14, 4.

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

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

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. Descriptive statistics

Panel (a) presents the market share of customer trades on MarketAxess in the US corporate bond market for different trade sizes 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" exceed \$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). Panel (b) provides descriptive statistics regarding size, bond age and time to maturity, rating, and trade direction for all trades that occur on TRACE (column 1) and for all trades that occur on MKTX (column 2) during our sample period. "Sell" inquiries represent customer sale requests. Our sample from MKTX covers all trading activity from January 3, 2017 to March 31, 2021.

Panel (a): MKTX market share of customer trades (%)			
	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

Panel (b): Descriptive statistics for MKTX and TRACE samples		
	TRACE (1)	MKTX (2)
HY	0.24	0.16
Sell	0.44	0.54
Size		
micro size (< \$100k)	0.62	0.51
odd lot (\$100k–1 million)	0.22	0.40
round lot (\$1–5 million)	0.12	0.08
block trade (> \$5 million)	0.04	0.01
Bond age distribution		
Average bond age	4.10	3.50
< 2 years	0.33	0.38
2–5 years	0.38	0.37
5–20 years	0.27	0.24
> 20 years	0.02	0.01
Bond time to maturity distribution		
Average maturity	7.45	9.00
< 2 years	0.18	0.09
2–5 years	0.31	0.33
5–20 years	0.40	0.42
> 20 years	0.09	0.14
Observations	27,866,961	5,630,755

Table 2. Responses to a traded and an untraded inquiry

Panel (a) provides dealers' disclosed responses to 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 customer accepted, the response status "Cover" flags the second best offer, and the response status "Missed" flags the rest of the responses that the customer rejected. Panel (b) provides dealer disclosed responses to 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 the customer studied in Table 2 (ID number 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 a 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), respectively, and trade probabilities for child orders at the first inquiry and after failing the first inquiry in columns (3) and (4), respectively. Credit ratings and trade size categories are defined in Table 1. "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 75$^{\text{th}}$ percentile of the distribution. To derive our measure of customer connectedness, we first regress the average number of dealer responses elicited by a particular customer on characteristics of the inquiry, and then rank customers into deciles based on residuals of this regression."Most connected" refers to the subsample in which the customer is in deciles 7–10, and "Least 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	0.8452	0.7429	0.3952
Sell	0.7518	0.8742	0.7853	0.4111
Buy	0.6590	0.8146	0.6984	0.3832
Investment grade (IG)	0.7321	0.8561	0.7691	0.3749
High yield (HY)	0.5898	0.7947	0.6207	0.4535
Micro size	0.7773	0.9042	0.8087	0.4976
Odd lot	0.6400	0.7945	0.6810	0.3519
Round lot	0.6678	0.7809	0.6918	0.2834
Block trade	0.6928	0.7828	0.7053	0.2588
High turnover	0.7103	0.8503	0.7478	0.4040
Low turnover	0.6923	0.8286	0.7265	0.3693
High amt outstanding	0.7621	0.8831	0.7932	0.4334
Low amt outstanding	0.6433	0.8009	0.6843	0.3653
Old	0.6620	0.8119	0.7013	0.3680
Not old	0.7204	0.8559	0.7563	0.4059
Most connected (deciles 7-10)	0.7956	0.9057	0.8193	0.4752
Least connected (deciles 1-6)	0.4075	0.6437	0.4544	0.3455

Table 6. Child order event statistics

This table presents summary statistics about child order events. By construction, the first event is always an inquiry on MKTX, either with or without trade. After a failed initial inquiry, each observation 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. The first row shows the probability of a failed and successful first inquiry on MKTX. The following rows provide 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. Unconditional Maximum Likelihood Estimation (MLE)

This table presents estimation results for the unconditional MLE, where the only control is a constant for event $k \in \{1, \ldots, 4\}$. 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 $N = 1,380,334$ observations.

Event	MKTX inq. w/o trade (1)	MKTX inq. w trade (2)	Voice trade (3)	Exit (4)
	-3.608***	-4.085***	-3.369***	-2.803***
	(4.915×10^{-6})	(7.692×10^{-6})	(4.612×10^{-6})	(3.183×10^{-6})

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

Table 8. Conditional Maximum Likelihood Estimation

This table presents our estimation results for the MLE conditional on trade characteristics and the number of failed inquiries in the child order. It shows how the intensities of each event, $\lambda(\theta_k' x)$, vary with trade characteristics. Columns (1) to (4) correspond to events $k \in \{1, ..., 4\}$. 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. "Ba1 to Caa3" takes the value of 1 if the bond's Moody's rating is between Ba1 and Caa3, and "Ca to C" is similarly defined. "High time-to-maturity" takes the value of 1 if the bond's time to maturity is above the sample median, and zero otherwise. "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. 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 $N = 1,380,334$ observations.

Event	MKTX inq. w/o trade (1)	MKTX inq. w trade (2)	Voice trade (3)	Exit (4)
(Intercept)	-3.930***	-3.370***	-3.180***	-2.870***
	(0.007)	(0.008)	(0.007)	(0.006)
Sell	0.107***	0.650***	0.175***	-0.007*
	(0.005)	(0.006)	(0.005)	(0.004)
Ba1 to Caa3	-0.081***	-0.117***	0.235***	-0.171***
	(0.006)	(0.008)	(0.006)	(0.005)
Ca to C	-0.113*	-0.324***	0.643***	-0.148***
	(0.064)	(0.086)	(0.051)	(0.052)
Covid	-0.167***	-0.573***	-0.264***	-0.182***
	(0.012)	(0.014)	(0.010)	(0.009)
Old	0.00334	-0.122***	-0.087***	0.029***
	(0.005)	(0.006)	(0.005)	(0.004)
Turnover below median	0.009*	-0.102***	-0.003	0.119***
	(0.005)	(0.007)	(0.006)	(0.004)
High time-to-maturity	-0.026***	0.041***	-0.111***	0.099***
	(0.005)	(0.006)	(0.005)	(0.004)
Low amt outstanding	0.143***	-0.339***	-0.343***	0.126***
	(0.005)	(0.006)	(0.005)	(0.004)
Micro size	0.052***	0.191***	0.270***	-0.221***
	(0.005)	(0.006)	(0.005)	(0.004)
Round lot	-0.283***	-0.582***	-0.007	0.236***
	(0.010)	(0.012)	(0.009)	(0.007)
Block trade	-0.446***	-1.32***	0.039	0.415***
	(0.032)	(0.046)	(0.031)	(0.024)
Connected decile < 7	0.047***	-1.660***	-0.042***	0.197***
	(0.006)	(0.008)	(0.006)	(0.005)
Connected decile 7	-0.082***	-0.580***	0.409***	0.068***

Table 8 continued

Event	MKTX inq. w/o trade (1)	MKTX inq. w trade (2)	Voice trade (3)	Exit (4)
	(0.010)	(0.013)	(0.010)	(0.009)
Connected decile 8	-0.170***	-0.880***	0.193***	0.122***
	(0.009)	(0.011)	(0.008)	(0.007)
Connected decile 9	-0.082***	-0.166***	-0.156***	-0.008*
	(0.007)	(0.007)	(0.007)	(0.006)
Failed 2	0.474***	-0.149***	-0.711***	-0.399***
	(0.006)	(0.008)	(0.007)	(0.005)
Failed 3	0.860***	-0.141***	-0.972***	-0.653***
	(0.008)	(0.014)	(0.013)	(0.009)
Failed 4	1.030***	-0.151***	-1.160***	-0.886***
	(0.011)	(0.020)	(0.021)	(0.015)
Failed 5	1.160***	-0.135***	-1.180***	-1.050***
	(0.014)	(0.029)	(0.029)	(0.021)
Failed 6	1.280***	-0.227***	-1.350***	-1.160***
	(0.016)	(0.040)	(0.042)	(0.029)
Failed 7	1.370***	-0.307***	-1.390***	-1.300***
	(0.020)	(0.053)	(0.055)	(0.039)
Failed 8	1.390***	-0.247***	-1.380***	-1.500***
	(0.022)	(0.063)	(0.066)	(0.051)
Failed 9	1.490***	-0.287***	-1.690***	-1.590***
	(0.027)	(0.076)	(0.093)	(0.065)
Failed 10	1.680***	-0.332***	-1.610***	-1.470***
	(0.015)	(0.046)	(0.051)	(0.035)

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

Table 9. Markup and trade probability for the first inquiry in child orders
This table reports the distribution of markups across child orders for the first inquiry, i.e., the distribution of $M_{i,1}$ defined in equation (3) across $i \in \{1, ..., 10\}$. We sort child orders into deciles based on $M_{i,1}$ and report the average markup and trade probability for each decile. The sample includes the first inquiries from all child orders. To mitigate the impact of extreme outliers, we winsorize markup at the 5% level.

Markup decile	Average markup (bps)	Trade probability
1	-22.61	0.851
2	-11.21	0.852
3	-6.21	0.845
4	-3.02	0.831
5	-0.46	0.812
6	2.07	0.788
7	5.20	0.758
8	9.86	0.720
9	18.10	0.668
10	41.24	0.568

Table 10. Price improvement relative to the first inquiry

This table reports the estimates corresponding to the regressions in equation (4) (columns 1 and 2) and in equation (5) (columns 3 and 4), where the dependent variable is the inquiry markup/markdown, $M_{i,j}$, defined in equation (3). "Inquiry j" takes the value of 1 if it is the jth inquiry in the child order and zero otherwise. "Inq. is traded" takes a value of 1 if the inquiry is traded on MKTX and zero otherwise. In columns (1) and (2), we include controls for trade direction (buy vs. sell), trade size (micro, odd lot, round lot, and block), and bond rating (IG vs. HY), described in Tables 5 and 8. The sample excludes inquiries that are submitted by dealers. Clustered standard errors at the customer level are shown in parentheses.

Dependent Variable:	inquiry markup (bps)			
Model:	(1)	(2)	(3)	(4)
Variables				
Inquiry 2	8.477***	-2.705***	17.04***	4.604***
	(0.9582)	(0.4107)	(2.142)	(0.6372)
Inquiry 3	12.21***	-2.542***	18.28***	7.512***
	(1.828)	(0.5892)	(2.876)	(1.249)
Inquiry 4	13.44***	-2.153***	18.58***	9.530***
	(2.931)	(0.8233)	(3.896)	(2.020)
Inquiry 5	14.79***	-0.6207	18.79***	11.34***
	(3.792)	(1.245)	(4.673)	(2.659)
Inquiry 6	13.75***	-0.8237	16.41***	11.35***
	(4.370)	(1.547)	(5.076)	(3.005)
Inquiry 7	12.09**	-0.4301	15.13***	12.83***
	(5.272)	(2.355)	(5.812)	(3.893)
Inquiry 8	11.51**	0.7160	13.14**	13.44***
	(5.651)	(2.906)	(5.812)	(4.367)
Inquiry 9	10.29*	0.4257	12.56**	14.28***
	(5.676)	(3.399)	(6.108)	(5.240)
Inquiry $\geq$ 10	8.404**	1.139	11.14***	16.56***
	(3.598)	(2.898)	(3.856)	(5.027)
Inquiry 2 × Inq. is traded			-18.42***	-15.83***
			(1.527)	(0.7134)
Inquiry 3 × Inq. is traded			-17.86***	-19.53***
			(1.913)	(1.003)
Inquiry 4 × Inq. is traded			-19.01***	-22.88***
			(2.427)	(1.397)
Inquiry 5 × Inq. is traded			-17.00***	-20.86***
			(3.034)	(1.714)
Inquiry 6 × Inq. is traded			-11.14***	-18.65***
			(3.340)	(1.670)
Inquiry 7 × Inq. is traded			-15.66***	-24.68***
			(2.833)	(2.107)
Inquiry 8 × Inq. is traded			-6.196*	-18.29***
			(3.555)	(2.728)
Inquiry 9 × Inq. is traded			-10.68***	-21.42***
			(3.764)	(2.650)
Inquiry $\geq$ 10 × Inq. is traded			-18.54***	-25.36***
			(3.265)	(2.630)
Controls	Yes		Yes	
Fixed-effects				
customer	Yes		Yes	
bond	Yes		Yes	
day	Yes	Yes	Yes	Yes
child order		Yes		Yes
Fit statistics				
Observations	7,275,262	7,281,083	7,275,262	7,281,083
R^2	0.09781	0.94165	0.10218	0.94378
Within R^2	0.00984	0.00238	0.01464	0.03891

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

Table 11. The role of dealer inventories: Extensive margin

This table reports OLS and logit estimates for the regression in equation (7), where the dependent variable, $dealer\ responded_{c,i,d}$, is a binary indicator that equals 1 if dealer d responded to inquiry i of child order c and zero otherwise. The main independent variables are EI^{bond} and $EI^{\text{aggregate}}$, which measure cumulative excess inventory from equation (6) at the dealer-CUSIP and aggregate dealer levels, respectively. We also include interactions of these variables with an indicator for a customer-sell trade. The regression include indicators for inquiry numbers 1 to 10 in the child order. In columns (1) and (3) we include the bond and trade characteristics described in Tables 5 and 8. In columns (2) and (4), we control for unobserved child order characteristics by adding child order fixed effects. The sample includes all dealers with at least 5 cumulative responses over a 120-day window from the inquiry date. To mitigate the undue influence of outliers, we winsorize EI^{bond} and $EI^{\text{aggregate}}$ at the 1% level. Clustered standard errors at the dealer level are shown in parentheses.

Dependent Variable:	dealer responded			
Model:	(1)	(2)	(3)	(4)
	OLS	OLS	Logit	Logit
Variables				
EI^{bond}	0.0312***	0.0326***	0.1861***	0.1696***
	(0.0042)	(0.0033)	(0.0249)	(0.0177)
EI^{bond} × Sell	-0.0442***	-0.0454***	-0.2628***	-0.2339***
	(0.0065)	(0.0060)	(0.0368)	(0.0288)
$EI^{\text{aggregate}}$	-0.0019	0.0269*	0.0011	0.1418*
	(0.0033)	(0.0138)	(0.0208)	(0.0748)
$EI^{\text{aggregate}}$ × Sell	0.0037	0.0026	0.0085	0.0053
	(0.0023)	(0.0026)	(0.0152)	(0.0127)
Inquiry number controls	Yes	Yes	Yes	Yes
Controls	Yes		Yes	
Fixed-effects				
customer	Yes		Yes	
bond	Yes		Yes	
issuer	Yes		Yes	
dealer	Yes		Yes	
day	Yes		Yes	
child order		Yes		Yes
Fit statistics				
Observations	88,442,065	89,799,560	88,439,876	86,762,314
Squared Correlation	0.29720	0.17437	0.31845	0.15376
Pseudo R^2	0.25109	0.13693	0.26345	0.12296
BIC	93,338,117.5	232,889,084.8	87,583,556.4	218,007,904.5

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

Table 12. The role of dealer inventories: Intensive margin

This table presents estimates for the regression in equation (8), where the dependent variable is $M_{i,j,d}$, the markup/markdown offered by dealer d in response to inquiry j of child order i, defined in equation (2). The main independent variables are EI^{bond} and $EI^{\text{aggregate}}$, measuring cumulative excess inventory from equation (6) at the dealer-CUSIP and dealer levels, respectively. We also include interactions of these variables with an indicator for a customer-sell trade. In both columns, we include dummies for inquiries 1 to 10 in child orders. In column (1), we include controls for trade direction (buy vs. sell), trade size (micro, odd lot, round lot, and block), and bond rating (IG vs. HY), 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. To mitigate the undue influence of outliers, we winsorize EI^{bond} and $EI^{\text{aggregate}}$ at the 1% level. The sample includes child orders that have at least two inquiries and excludes inquiries submitted by dealers. Clustered standard errors at the dealer level are shown in parentheses.

Dependent Variable:	markup (bps)	
Model:	(1)	(2)
Variables		
EI^{bond}	-1.821***	-1.530***
	(0.1921)	(0.1631)
$EI^{\text{bond}} \times$ Sell	2.830***	2.613***
	(0.4299)	(0.3034)
$EI^{\text{aggregate}}$	0.3219	1.268**
	(0.2635)	(0.5558)
$EI^{\text{bond}} \times$ Sell	-0.3204	-0.5987
	(0.5393)	(0.4135)
Inquiry number controls	Yes	Yes
Controls	Yes	
Fixed-effects		
customer	Yes	
bond	Yes	
issuer	Yes	
dealer	Yes	
day	Yes	Yes
child order		Yes
Fit statistics		
Observations	3,164,852	3,165,865
R^2	0.14228	0.54548
Within R^2	0.00362	0.00438
Clustered (dealer) standard-errors in parentheses		
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*		

Appendix

A A McCall (1970) Model of Customer Search

A.1 The Model

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 below, in Appendix A.3, 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 $j \in \{0, 1, 2, \dots\}$ 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,

$$m^\star = \frac{c}{r + \gamma} - \frac{\lambda}{r + \gamma} \int_0^{m^\star} F(m)\, dm. \tag{A1}$$

The first term in Equation (A1), $c/(r + \gamma)$, 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 Arrival Rates and Estimation Biases

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, at any point in time, there is an inflow $d\phi(x)$ of type-x child orders in the market. Hence, the measure of type-x child orders with $n \geq 0$ failed inquiries satisfies the inflow-outflow equations:

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

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:

$$d\mu(x \mid x) = \pi_1(x)^n d\mu(x \mid 0) \text{ where } \pi_1(x) \equiv \frac{\lambda_e(x)\left(\sum_j q_j \left[1 - G_e(m^\star(x) \mid x)\right]^j\right)}{\lambda_e(x) + \lambda_v(x) G_v(m^\star(x) \mid x) + \gamma(x)}, \tag{A2}$$

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 (A2), 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 $\pi_1(x)$ 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.

We now establish this result more formally. In what follows, it will be useful to 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.

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) $H(x \mid n-1)$.*

Proof. 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.

A.3 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.

Internet Appendix

IA.1 Dealer excess inventory measurement

In this section, we provide a detailed explanation of how we construct the measure of dealers' inventory holdings introduced in Section 4.4. To measure inventory at the dealer level, we use the academic version of the TRACE corporate bond database, which contains anonymized dealer identifiers. These identifiers enable us to determine the dealer(s) involved in every transaction and to track the *changes* in dealers' inventory over time. However, since we do not know the initial level of inventory held by a dealer at issuance, we are unable to calculate the *level* of a dealer's inventory. Because of this limitation, we use changes in cumulative inventory—at the bond level and at the aggregate level—to create two different measures of each dealer's "excess inventory," or EI_d^ℓ, defined in (6) for $\ell \in \{\text{bond}, \text{aggregate}\}$.

As noted in the text, this measure of excess inventory requires calculating changes in a dealer's inventory between an initial date $\tau = 0$ and the current date $\tau = t$:

$$I_{d,t}^\ell - I_{d,0}^\ell = \sum_{\tau \in T_{d,t}^\ell} Q_{d,\tau}^\ell,$$

where $T_{d,t}^\ell$ denotes the set of trades executed by dealer d between $\tau = 0$ and $\tau = t$ for either the inquiry-specific bond (ℓ = bond) or all bonds (ℓ = aggregate), and $Q_{d,\tau}^\ell$ denotes the corresponding quantity traded, which is positive for dealer purchases and negative for dealer sells. Intuitively, at either the bond-level or the aggregate-level, EI_d^ℓ captures the dealer's current inventory position relative to their "typical" inventory holdings.

IA.1.1 Mapping Dealer Identifiers

After calculating the excess inventory for each dealer identified in academic TRACE, we must map them to the anonymized dealer identifiers in MKTX. We proceed as follows:

1. **Sample selection.** We first select the sample of customer inquiries that resulted in trade on MarketAxess. For this subsample, we collect the traded quantity, the trade side (buy or sell), the timestamp of the trade, and the anonymized identifier of the dealer counterparty.

2. **Trade matching procedure.** We then use trades to match identifiers on MKTX with those in TRACE. Specifically, for each trade in our MKTX subsample, we require an exact match on the CUSIP of the bond, the traded quantity, and the trade side. Among the TRACE records that satisfy these three criteria, we select the record that is closest in time, with a maximum tolerance of one minute difference between the trade time recorded in TRACE and the trade time recorded in MarketAxess; this stringent criteria on trade time reduces the likelihood of erroneous matches. We do not match on trade prices because a significant fraction of the trade prices on TRACE and MKTX are not identical.[1] Following this strict time matching approach, we can match 54% of the trades in MarketAxess to a record in academic TRACE.[2] The outcome of this matching exercise is a dataset where we observe the corresponding MKTX and TRACE anonymized identifiers for every matched trade.

3. **Dealer identifiers: frequency table and mapping procedure.** In many cases, the traded responses of a given anonymized entity in MKTX map to multiple anonymized entities in TRACE.[3] To determine the most likely corresponding entity, we create a frequency table for the TRACE dealer identifiers to which the MKTX counterparty has been matched with. Using this frequency table, we determine the TRACE identifier of a MKTX dealer by selecting the highest frequency in the table, subject to the following criteria:

   (a) The dealer has at least 100 traded responses in MKTX.

   (b) The relative frequency of the MKTX-TRACE map with the highest frequency is greater than 90%.

   The first criteria is a rule of thumb that has the purpose to ensure that we have enough data to be confident in the mapping. The second criteria limits our mappings to the set of dealers for which we have a high degree of confidence that our approach has correctly identified the appropriate trace match. Following this procedure, we succeeded in assigning an academic TRACE dealer identifier to 99% of the responses submitted on MarketAxess.

[1]We conjecture that some of the discrepancy between the trade prices reported in TRACE and MKTX might be attributed to the inclusion of the MKTX platform fee in the reported price.

[2]Extending the tolerance window to 5 minutes increases the fraction of matched trades to around 80%. Using this extended window has no impact on the outcome of the procedure.

[3]While slight difference in reporting times between the two data sources undoubtedly introduce mismatches, our analysis suggests that some of the MKTX entities genuinely report under multiple TRACE entities.

IA.1.2 Extensive margin calculations

To estimate the propensity of a given dealer to respond to an inquiry, we face an important data limitation: while we observe the number of disclosed dealers an inquiry reached, we do not observe the identity of all the contacted disclosed dealers. We only observe the identity of the dealers that respond to an inquiry. We circumvent this data limitation by inferring the identities of the disclosed dealers that were contacted based on the following two-step approach.

First, we determine the list of dealers that have replied to the customer's inquiries in the past 120 days. If a dealer has not replied to a customer's inquiries over the last few months, he is unlikely to be in the disclosed list of the customer. Second, among the list of dealers who replied to the customer in the past 120 days, we select those who have traded in the bond at least five times over the same period. If a dealer is not actively trading a given CUSIP, a customer is unlikely to send him the request. Moreover, to estimate our extensive margin regression (7) we must observe some amount of trading by a dealer to be able to calculate his excess inventory.

To evaluate the accuracy of this approach, we calculate the frequency at which customer inquiries receive disclosed responses from dealers that are not in the customer's list of inferred disclosed dealers. If our approach correctly identifies the list of a customer's disclosed dealers, we would expect the frequency to be very low. This is indeed what we find: only 0.17% of disclosed responses are from dealers that are not in the list of inferred disclosed dealers that we constructed for the customer submitting the inquiry.

It is more difficult to test the prevalence of the other type of error (when a dealer in the inferred dealer list does not receive the inquiry submitted.) However, this type of measurement error clearly biases our estimates against finding an effect of dealer inventory on their propensity to respond to an inquiry.

IA.2 Additional Figures and Tables

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. Markup for the first inquiry and trade probability
This table presents OLS and logit regression estimates. The dependent variable is an indicator for whether the inquiry is traded on MKTX and the main dependent variable is inquiry markup for the first inquiry in child orders, defined in Equation (3) and standardized to have 0 means and unit standard deviation. In columns (1) and (2) we include the following dummies for customer, bond, and trade characteristics described in Tables 5 and 8: sell, HY, micro size, odd lot, round lot, block trade, low amt outstanding, old, connected decile 7, connected decile 8, connected decile 9, and connected decile below 7. The sample include first inquiries in all child orders. Clustered standard errors at the customer level are shown in parentheses.

Dependent Variable:	inquiry is traded			
Model:	(1)	(2)	(3)	(4)
	OLS	Logit	OLS	Logit
Variables				
Markup, scaled (bps)	-0.0867***	-0.5940***	-0.0874***	-0.6747***
	(0.0051)	(0.0250)	(0.0041)	(0.0220)
Customer and bond controls	Yes	Yes		
Fixed-effects				
day	Yes	Yes	Yes	Yes
customer			Yes	Yes
bond			Yes	Yes
Fit statistics				
Observations	5,683,376	5,683,376	6,830,437	6,828,165
Squared Correlation	0.19674	0.20165	0.37834	0.38881
Pseudo R^2	0.25200	0.20065	0.42868	0.35228
BIC	3,710,776.8	4,125,775.7	4,559,778.4	4,998,758.0

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

Table IA.5. 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 controls for customer connectedness, trade direction (buy vs. sell), trade size (micro, odd lot, round lot, and block), bond age, Covid (inquiries submitted in Mar 2020), bond rating (IG vs. HY), and amount outstanding 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 r egression. 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*