Portfolio Trading in Corporate Bond Markets

Jeffrey Meli *†

Zornitsa Todorova*

Barclays, US

Barclays, UK

2nd March 2023

Abstract

Portfolio trading, a recent innovation in the corporate bond market, involves trading a basket of bonds as a single piece of risk with a single market-maker. Using a proprietary dataset of portfolio inquiries, we develop an algorithm to identify portfolio trades in TRACE. We estimate that portfolio trading has increased from 0% of total investment grade corporate bond volumes at the beginning of 2018 to over 7% in 2021. Portfolios are designed to generate liquidity in illiquid bonds, and the protocol is remarkably cost effective. We show that portfolio trading reduces execution costs by over 40%, with the largest benefits accruing to the least liquid bonds. We provide evidence that spill-overs from the ETF ecosystem allow market-makers to offload the inventory of illiquid bonds which accumulates as a result of portfolio trading and to both price and hedge portfolio trades.

JEL Codes: C55, G12, G14

Keywords: portfolio trades, corporate bonds, transaction costs, ETFs

† Corresponding author: Jeffrey Meli, Tel: +1 212 412 2127, email: *jeff.meli@barclays.com* Barclays, 745 7th Ave, New York, NY 10019

1. Introduction

In this article we analyse portfolio trading, the latest innovation in the corporate bond market. In this new trading protocol, an investor bundles a set of individual corporate bonds into one basket and asks a market-maker or market-makers to quote the entire basket as a single piece of risk, instead of seeking quotes on each bond individually. The investor then executes the entire basket with the market-maker that provided the best price. This differs from the standard request-for-quote (RFQ) trades, which are executed on an individual bond basis.

We make both methodological and analytical contributions towards a better understanding of portfolio trades. Our main methodological contribution is to build a comprehensive database of portfolio trades by applying a machine learning clustering algorithm to the Trade Reporting and Compliance Engine (TRACE) data. The algorithm is built using insights from a proprietary dataset of portfolio trade inquiries received by a large market-maker. Our data show that portfolio trading experienced rapid growth, from 0% of total corporate bond volumes in 2018 to 7% in 2021. We use the dataset of PTs identified by our algorithm to better understand this growth, in terms of both the motivation for using this new protocol and its effectiveness. First, we show that the main motivation for executing PTs is to generate liquidity in illiquid bonds, rather than market positioning or speed of execution. Second, we show that PTs are remarkably effective at generating liquidity; transacting via PTs reduces realized transaction costs by more than 40%, with the bulk of the gains accruing to the least liquid bonds. Finally, we attribute these reduced transaction costs to spill-overs from the ETF ecosystem, which provides both demand for less liquid bonds through the create and redeem (C/R) process and price transparency for diversified corporate bond risk.

The link between the effectiveness of portfolio trading and the ETF ecosystem is the key insight of our paper. Several prior studies have investigated the direct effect of ETFs on the risk, return, and liquidity of the underlying assets (e.g. Ben-David, Franzoni, and Moussawi (2018), Da and Shive (2018), Holden and Nam (2019) among others). Our main contribution to the literature is to show that the implications of ETFs can extend beyond their direct effects. For corporate bonds, synergies with the ETF ecosystem have enabled the development of an entirely new trading protocol, one so effective that it reduces transaction costs by 40%. Portfolio trading is wholly reliant on the existence of ETFs, despite the fact that the investors utilizing the protocol need not ever buy or sell an ETF directly. Improved corporate bond liquidity can have its own implications, such as for example lower yields. Our results suggest that understanding the overall effect of ETFs on the underlying markets requires an assessment of these indirect channels as well as the direct channels that have been the subject of the existing literature.

Building a database of portfolio trades in TRACE is a challenging task because they are not currently flagged in the TRACE feed. The reporting requirements will change starting in May 2023, reflecting the growing importance of this new trade protocol.[1] Therefore, a necessary first step in the analysis is to distinguish PTs from RFQ trades. To do so, we use a proprietary database of PT inquiries in investment grade (IG) corporate bonds received by a large market-maker. We then match them to the TRACE feed to find a verified set of PTs that actually traded. Because investors are conscious of revealing potentially sensitive information to many counterparties at a time, they typically send the inquiry to a few market-makers only. This implies that the matched inquiries are not a complete set of all portfolios that actually traded. To overcome this limitation, we use those verified PTs to develop a machine learning clustering algorithm, which can identify additional portfolio trades that are not already part of our inquiry database. Using this methodology, we build a fulsome dataset of portfolio trades

[1] The reporting rule will change on the 15th May 2023, as described in FINRA's *Regulatory Notice 22-12*.

spanning more than 12,000 unique IG PTs and c.1 million bond-PT transactions. We perform a number of validation checks on this dataset to ensure that the algorithm identifies actual portfolio trades.

Using this novel database, we show that portfolio trading has grown rapidly, from virtually no transactions in 2018 to 7% of the total transaction volumes in 2021. This rise mirrors the increase over time in the number of inquiries in our database and demonstrates why the activity level is high enough to justify the addition of a PT flag in the TRACE feed.

By analysing and comparing the characteristics of the bonds included in portfolio trades and in standard RFQ trades, we conclude that the main motivation for executing via portfolio trades is to improve the ability to trade illiquid bonds. First, the portfolios traded tend to have lower liquidity than the trades done in the standard RFQ format. Second, this reduced liquidity is not uniform across the typical portfolio. Instead, portfolios combine some very illiquid bonds with other highly liquid bonds. We interpret this as an attempt to "crowd-source" liquidity in illiquid bonds whereby the overall transaction cost is reduced by bundling them with more easily traded securities. Finally, even in portfolios that appear to target specific market segments (such as the long end of the credit curve) most of the portfolios heavily feature illiquid bonds.

We find that portfolio trading significantly reduces transaction costs. We employ a rigorous panel regression specification, where we compare the difference in the realized transaction cost of portfolio trades and standard RFQs for the same bond on the same day. We control for trade-level and time-varying bond-level characteristics, and include both bond and day fixed effects. In aggregate, transacting via a portfolio trade reduces transaction costs by over 40% versus the standard RFQ protocol.

The reduction in transaction costs is not uniformly distributed across bonds. The greatest benefit accrues to the least liquid bonds, whereas the benefit is very limited for the most liquid bonds. All else equal, we find that portfolio trading is 30% more cost effective for illiquid than for liquid bonds. This result holds for a diverse set of liquidity measures such as transaction costs, trade volume, price impact, autocorrelation in returns and bond age.

That PTs reduce transaction costs easily explains their popularity from the perspective of investors, but leaves unanswered the question of why PTs reduce transaction costs. Phrased differently, why are market-makers willing to execute them at such low transaction costs? We propose that spill-overs from the ETF ecosystem drive the reduction in PT transaction costs. These spill-overs come from two features of ETFs. First, ETFs routinely transact in bonds via the create and redeem process, generating activity in bonds that they own. Second, corporate bond ETFs are extremely liquid (both in absolute terms and relative to the underlying bonds (Meli and Todorova, (2022)), and trade actively in the secondary market. PTs benefit to an equal degree from both of these features. First, ETFs provide an outlet to offload the inventory of bonds accumulated from portfolio trades via the C/R process. This is particularly important for the illiquid bonds included in a PT, which might otherwise be difficult or costly to trade, and for which the benefits of transacting via a PT are particularly strong. Second, ETFs provide a useful real-time tool to transparently price diversified baskets of corporate bonds and an efficient hedging tool for market-makers when purchasing or selling bonds via PTs. A diversified basket of corporate bonds can be hedged using an ETF, whereas hedging a single name position with an offsetting position in an ETF incurs substantial basis risk.

We find strong empirical evidence to support both channels. There is a high degree of overlap between portfolio trades and ETFs; on average 60% of the line items included in PTs are also owned by the largest IG ETFs, whereas these ETFs only own about 30% of the bonds in the broad IG corporate bond index. In panel regressions, we explore the transaction costs of trades that are "right way" vis-à-vis the ETFs, meaning that investors are selling bonds when ETFs are net creating shares, and vice-versa. The "right way" effect is 60% higher for PTs than for comparable RFQ trades. Moreover, the cost reductions are the highest for the least liquid bonds, which highlights the importance of aligning PTs with the direction of the ETF C/R flows for illiquid bonds. The economic intuition behind this result is that market-makers put better prices on illiquid bonds included in PTs that can be packaged with other debt and offloaded via the C/R process in a timely manner.

In principle, the ETF C/R channel could apply equally to both PTs and RFQs, yet we find that being "right way" is more important for PTs. Using intra-day pricing data, we show that ETFs provide benefits of price transparency, price discovery and risk hedging, which only apply when the bonds are traded in portfolio form, and not individually. PT execution costs are strongly positively correlated to ETF deviations from NAV. For example, investors selling bonds in PTs incur transaction costs that are c.0.6 basis points lower when the ETF is trading 1 basis point above NAV (and c.0.6 basis points higher transaction costs when the ETF trades 1 basis point below NAV). In contrast, the transaction costs for trades in the same bonds, on the same day, but traded in RFQs, have no statistically significant correlation with the ETF deviations from NAV, and instead depend entirely on bond characteristics.

Relationship to prior literature

Our analysis contributes to several areas of the existing literature. First, we contribute to the literature that studies how the supply of and demand for corporate bond liquidity has evolved since the global financial crisis (GFC). Several papers have shown that corporate bond liquidity has deteriorated in the aftermath of the GFC (e.g. Dick-Nielsen, Feldhütter, and Lando (2012); (Friewald, Jankowitsch, & Subrahmanyam (2012); (Bessembinder, Jacobsen, Maxwell, & Venkataraman, (2018)). Against this backdrop, a large body of work has investigated how the supply of liquidity provided by market-makers has changed with market conditions, regulations and trading protocols (e.g. Goldstein and Hotchkiss (2020), Goldberg and Nozawa (2020) and Carapella and Monnet (2020) among others). For example, as market-makers became less willing to hold inventory, more trades were done on an agented basis (meaning market-makers line up the other side of the trade before executing), which involves a trade-off between transaction costs and immediacy and certainty of execution. Other research has instead focused on how investors adapt to lower liquidity. Jiang, Li, and Wang (2021) demonstrate that open-end corporate bond funds dynamically manage liquidity to meet investor redemptions; Meli and Todorova (2022) show that high yield mutual funds use ETFs to manage liquidity, which results in an aggregate decline in high yield bond liquidity as investors substitute trading in ETFs for trading in the underlying bonds.

Our analysis documents the next stage in the development of new trading protocols and the management of liquidity needs, as investors take advantage of new developments in the market to innovate further on trading protocols and mitigate the effect of reduced liquidity. In this paper, we propose that the rise of corporate bond ETFs naturally generates demand for transactions in bonds that are owned by the ETFs. The majority of ETF bonds are liquid and trade frequently. However, liquidity is typically highest immediately after issuance, and tends to decline steeply as bonds age. ETFs have demand to transact in these older, less liquid bonds in order to limit deviations from their benchmarks that arise as their AUM changes. We show that portfolio trading effectively piggy backs on this demand to provide liquidity in illiquid bonds.

Another area of the literature that we contribute to relates to the implications of ETFs for the underlying financial markets. For equity ETFs, the literature shows that ETFs have a positive effect on volatility (Ben-David, Franzoni, and Moussawi (2018)), return co-movement (Da and Shive (2018)) and liquidity co-movement (Agarwal et al. (2018)). For bond ETFs, several studies show that ETFs lead to better liquidity (e.g. Holden and Nam (2019), Ye (2019), Marta (2020), Meli and Todorova (2022)) and better price discovery (Choi, Kronlund and Oh (2022)), but could weaken bond price informativeness (Rhodes and Mason (2022)) and increase bond fragility (Dannhauser and Hoseinzade (2022)).

The main insight of our analysis is that ETFs are not just an innovation unto themselves, but that they can be the source of further innovations and spill-over effects. Our work complements two ongoing studies that examine spill-overs from corporate bond ETFs. Shim and Todorov (2021) document that ETF C/R baskets are fractional and discuss implications for ETF premiums and discounts. Koont et al. (2022) show that basket inclusion generates additional trading activity, which improves the liquidity of the bonds in the ETF baskets. Our paper proposes a novel channel through which ETFs can impact corporate bond liquidity. We show that investors and market-makers have taken advantage of the real-time price transparency that secondary trading in ETFs provides and the need to execute transactions in specific underlying bonds driven by the create and redeem process to create a new trading protocol. Our analysis shows that portfolio trading is an incredibly cost effective way to generate liquidity in illiquid bonds.

2. Data and Variables Definitions

2.1 Portfolio Trade vs. RFQ Protocol

In the standard RFQ protocol, an investor requests a quote on a bond from one or several market-makers, and typically transacts (if at all) with the market-maker that provided the best price. Investors at times make RFQs for a large number of bonds at once.[2] The responses are evaluated on a bond-by-bond basis; the investor executes each line item individually with the

[2] These lists are known as “BWICs” (bids wanted in competition) or “OWICs” (offers wanted in competition).

market-maker that provided the best quote for that bond, with no expectation that the transactions will be pooled or bundled.

In a portfolio trade investors ask market-makers to price an entire portfolio as a single piece of risk. If the investor agrees to the price, the portfolio trade is then executed in its entirety with a single market-maker. Like with an RFQ, an investor can request a quote on a portfolio from one or several market-makers. Although a single price is agreed to for the entire portfolio, each individual line item is still subject to the TRACE (Trade Reporting and Compliance Engine) reporting rules. The prices of the individual line items reported to TRACE (weighted by their respective notionals) must sum to the quoted price of the portfolio. There are several reasons why both investors and market-makers ensure that the prices reported to TRACE are accurate at the bond level (i.e., that the portfolio price is correctly apportioned across the individual line items). For example, investors have best execution requirements that apply at the bond, and not at the portfolio, level. An investor who bought a portfolio where some bonds were priced too richly would attract scrutiny, regardless of whether other bonds in the portfolio were priced cheaply. Similarly, market-makers are subject to fair dealing requirements. A dealer that reported a bond purchase to TRACE at an artificially low price could be accused of excessive mark-up if it then sold the bond at the correct price to a different investor. Therefore, the transaction costs of PTs can, in principle, be compared to those of RFQs at the bond level. That said, PTs are not currently flagged in TRACE, which presents a practical challenge to any such comparison to RFQs.

2.2 Portfolio Trade Inquiry

To identify PTs, our first step is to collect two proprietary datasets of investment grade (IG) investor portfolio inquiries received by the Barclays corporate bond trading desk.

The primary data set we use covers the period 1st October 2018- 31st December 2021. Our sample contains all inquiries received by the trading desk, regardless of whether or not

Barclays executed the trade. The inquiry dataset spans c. 3,000 investor inquiries that contain c.22,000 unique investment grade bonds (each bond identified by a CUSIP). For each inquiry we obtain the date when it was received (but not the exact time stamp), and the CUSIP, notional, and direction (buy or sell) for each line item in the portfolio. Panel A of ***Table I*** gives an example of such a typical inquiry.[3]

The number of portfolio trade inquiries in the dataset grew significantly over the sample period, from virtually zero in 2018 to more than 2,000 inquiries and $175 billion in volume in 2021 (***Figure 1***). While we believe that Barclays has a large enough market share such that the sample of inquiries we receive is representative in terms of line items, volumes, and execution times, the inquiry dataset is not a full accounting of all PT inquiry in the market. No single market-maker has access to the complete set of inquiries because institutional investors balance the potential for price improvement from submitting their prospective portfolio trade to many counterparties against the risk of revealing market-moving information. This motivates the need to construct a more comprehensive database of portfolio trades. Hence, in Section 3, we use the proprietary inquiries database to develop an algorithm which can identify portfolio trades which are not included in our inquiry data.

The secondary dataset of PT inquiries covers the period September 1st 2022 – 26th January 2023. Although considerably smaller in size, the advantage of this sample compared to the primary dataset is that the it also includes intraday pricing information (which is the lacking in the earlier dataset): the exact time stamp (rather than just the date) of the inquiry, along with the Bloomberg bid/offer bond quotes for each line item at that time. We pair this

---

[3] The dataset also contains a flag if the portfolio is "custom" or "in-competition". We discard the custom portfolios, which make up about 10% of the sample, because they are designed by the market maker to achieve a particular investment objective for the investor (e.g., the investor wants to buy $150 million BBB-rated, 12+ maturity debt). Since it is possible that the line items in these custom inquiries are influenced by the market-maker's existing inventory and risk appetite, they might not be representative of the market.

with data on the deviations from NAV of LQD, the largest IG ETF, at the exact times of the inquiry. Panel B of ***Table I*** gives a sample of this data; we use this intra-day data when assessing the linkages of PTs to ETFs in Section 6.

2.3 Bond Sample and Liquidity Measures

Bond Sample

We obtain transaction-level corporate bond trading data from the standard version of TRACE, which caps trade sizes at $5 million for IG bonds, for the period 1st October 2018-31st December 2021.[4] We follow the approach by Dick-Nielsen ((2009), (2014)) to remove double counting, corrections, reversals and cancellations from TRACE. We then augment the cleaned daily TRACE data with bond-level characteristics from Bloomberg (spread, maturity, time since issuance, numeric rating, amount outstanding, issue size, sector classifications and call types), computed at the beginning of each month. The Bloomberg data covers dollar-denominated bonds belonging to major bond indices (e.g. Bloomberg Investment Grade Corporate Bond Index). We drop bonds with incomplete or missing data. The resulting bond data contains records for 97% of the line items in the portfolio inquiry dataset.

Liquidity Measures

We compute five liquidity metrics at the bond-month level: Liquidity Cost Score (LCS), bond age, Trade Efficiency Score (TES), Price Impact, and Roll's measure. LCS is a commercially available measure of transaction cost computed using quotes from the Barclays trading desk. It follows the methodology by Konstantinovsky, Yuen Ng, and Phelps (2016). LCS measures the transaction cost for an institutional-size round-lot trade, expressed as a percentage of the bond's price (hence higher LCS signifies lower liquidity). We also use

[4] The enhanced TRACE data disseminates uncapped trade sizes, but is only available to us with a 6-month delay. This is why we prefer to work with the standard data file, which is available immediately. Since individual line items in a portfolio trade rarely exceed the cap, working with the standard TRACE data instead of the enhanced version does not have a material impact on our analysis.

bond age as a proxy for liquidity, the intuition being that bonds are most liquid shortly after issuance, and as bonds age, their liquidity tends to decrease.

The other measures are computed using transaction data from TRACE.[5] TES blends transaction costs and trading volume into a single relative trade score, reflecting both the cost and the flow. To calculate TES, we assign each bond to a monthly LCS quintile and a monthly trading volume decile. Then, the sum of these quantiles (ranging from 2 to 15) is mapped to a TES ranking from 1 to 10, where a lower TES corresponds to better liquidity. We define Amihud's (2002) daily measure of price impact as the volume-weighted absolute daily return. Then, to convert to a monthly frequency, we use the median value of the daily price impact in that month. Finally, in the spirit of Roll (1984), we compute the first-order auto covariance using all transaction level price changes within a given month.

There are advantages and disadvantages of both quotes-based and trade-based liquidity measures (Schestag, Schuster, & Uhrig-Homburg (2016)). The advantage of quotes-based measures is that they are not limited to realized transactions only. The concern, however, the quotes from a single market-maker on an individual bond could reflect the inventory and risk appetite of that particular market-maker. This could introduce noise but not bias, as any positioning or risk tolerance will average out over time and across bonds. The advantage of liquidity measures computed from TRACE data is that they include all trades and are not likely to be influenced by a single market-maker's inventory position. On the other hand, the typical concern about TRACE-based liquidity measures is that they are computed from realized transactions only, which may present a distorted picture of liquidity. For example, trading by "appointment", in which a market-maker only executes a trade after both sides have been identified, incurs lower transaction costs than pure principle based trading, in

[5] For other liquidity measures refer to work by (e.g. Puh (2009), Feldhütter (2012), Dick-Nielsen, Feldhütter, and Lando (2012), Corwin and Schultz (2012)), and frequency of trading (also known as zeros) (Lesmond, Ogden, and Trzcinka (1999)).

exchange for reduced immediacy and certainty of execution, implying that the realized transaction costs are below the true transaction costs. Combining quotes-based and transaction-based measures helps overcome these concerns and paints a more holistic picture of liquidity in the corporate bond market.

2.3 Daily ETF C/R Baskets

We construct daily ETF C/R baskets for LQD, the largest IG ETF. Following the methodology by Shim and Todorov (2021) and Koont et al. (2022), we impute LQD's realized creation and redemption baskets from daily changes in holdings on days with C/R activity. Daily ETF holdings are publicly available and can be downloaded from the iShares website.[6] We identify create (redeem) days as those days on which there was a positive (negative) change in the number of LQD shares. We then use daily changes in the number of bonds held to infer the composition of the average LQD basket on each day.

It is possible that there are redeem baskets on days with net creations and create baskets on days with net redemptions, and that different authorized particpants (APs) negotiate different baskets with an ETF on the same day. Therefore, our imputed baskets are best interpreted as the average net basket for LQD on each given day. As Koont et al. (2022) discuss, this methodology would mis-characterize trades executed by ETFs directly in the secondary market. However, such trades are uncommon because, differently to in-kind C/R transactions, they incur tax liabilities for investors. We have verified the average (monthly) correlation between actual LQD flows reported by Bloomberg and the flows implied by our methodology is close to 0.80 (***Figure A2. 1***).

---

[6] https://www.ishares.com/us/products/239566/ishares-iboxx-investment-grade-corporate-bond-etf

3. Constructing a Database of Portfolio Trades

Our first contribution is a methodological one; we use our proprietary dataset of portfolio inquiries to develop a machine learning algorithm to identify a fulsome set of portfolio trades in the TRACE database. The dataset we construct includes more than 12,000 unique IG PTs and c. 1 million bond-PT transactions.

3.1 Methodology

In developing our methodology, we seek to balance classification error against the ability to find as many portfolio trades as possible. We proceed in four steps (***Figure 2***). [7]

In the first step, we match the portfolio inquiries to TRACE to identify which inquiries actually traded. Generally, we either don't find the inquiry in TRACE at all or we find it in full or very nearly so. For example, we find 68% of the inquiries in full. This "take-it-or-leave-it" nature of portfolio trades works to our advantage because it allows us to obtain a clean set of traded inquiries, without worrying that the line items we have not been able to match (for whatever reason) could introduce a large degree of noise in our model. We then analyse the matched inquiries and construct the blueprint of the typical portfolio trade in TRACE in terms of the distribution of execution time stamps, number of line items, volumes, average trade sizes etc.

We find that the trades associated with an individual portfolio trade appear like spikes or clusters in the TRACE data, with the same or very similar execution time stamps. Hence, in the second step, we run a machine learning clustering algorithm on the TRACE data. This clustering algorithm classifies the TRACE trades into two types of trades: clusters of "candidate" portfolio trades executed within a window of a few seconds, and all other (non-

[7] The Data Appendix contains more details on each of these steps. The Python code we used to identify portfolio trades is available upon request. Although we restrict our analysis to the IG market only, the code is designed in a way that allows researchers to construct a HY portfolio trades database as well.

portfolio) trades. The time we allow to elapse between the line items in each candidate portfolio trade is a conservative estimate of the patterns we see in the matched inquiry database.

Third, we re-cluster portfolio trades in order to group together "legs" of the same portfolio trade. This is motivated by the fact that in some cases portfolio trades are reported in TRACE in blocks, separated by a few minutes. This is most common for the different legs of long-short portfolios; for example, the long leg may be reported at 11:45:10 and the short leg at 11:46:50. If we increased the clustering window in Step 2 the model will correctly identify that both legs belong to the same portfolio, but at the cost of identifying many false positives, which are trades that simply happened to be executed between the times of 11:45:10 and 11:46:50. It is worth noting that if we were only interested in an aggregate estimate of the PT volumes, re-clustering is not necessary (the estimate of total size of PTs is not affected if $100 million of PT volumes were generated by one or two different portfolio trades). However, analysis of the linkages between PTs and ETFs requires knowing more precisely the composition of each portfolio.

Finally, we apply a series of filters to the data to convert the candidate portfolio clusters into actual portfolio trades. The two most restrictive filters are the exclusion of candidate portfolio trades with fewer than 25 line items and of candidate portfolio trades executed around popular delayed spotting times.

Although c.10% of our inquiries have fewer than 25 line items, we believe many of these are not strictly speaking PTs and are not representative of how PTs are actually priced.[8] This

[8] For example, if an investor mistakenly requests a price for a list of 10 bonds via the PT protocol instead of the RFQ protocol, we would capture the list in the inquiries database, and potentially see the line items subsequently printed in TRACE. However, in reality, this was executed in an RFQ. Alternatively, when investors first use the PT protocol, they typically request PT quotes for a smaller basket of bonds; as they get more comfortable with the process, they increase the size of the PT basket. However, the execution quality am investor would get for a small PT with a limited number of CUSIPS would be very different from the execution of a larger PT.

filter is also in part informed by FINRA's definition of portfolio trades, according to which a portfolio trade involves at least 10 unique corporate bonds.[9] We use a stricter definition and apply additional requirements for the minimum trade volume and average trade side in order to ensure that we capture institutional-size transactions and limit classification error.

IG bond volumes in TRACE exhibit sharp daily spikes around known times, which represent delayed treasury spotting.[10] Transactions in IG corporate bonds are often made by counterparties agreeing on a spread to a benchmark Treasury. The actual dollar price of the trade is computed at a later point in time using the previously agreed upon spread. For example, the counterparties agree to a spread at 13:30, but the price is calculated and reported to TRACE at 15:00. Delayed spotting allows investors and market makers to concentrate (and net) their Treasury hedging, instead of having to do multiple hedges throughout the day. While we know from our inquiries that some portfolio trades are reported around spotting times, the sheer amount of trades that are concentrated around these times makes it impossible to accurately separate portfolio trades from delayed spot RFQ trades. To ensure that our comparison between PTs and RFQs is always meaningful, we also drop RFQ trades executed around spotting times from our regression sample.[11]

3.2 Summary Statistics and Algorithm Validation

The resulting portfolio trades database that we construct contains 12,107 unique IG portfolio trades and c. 1 million bond-portfolio observations, totalling $696 billion of executed bond volumes (Panel A, ***Table II***).

[9] FINRA's *Regulatory Notice 22-12*

[10] We estimate that in 2021 between 7% - 12% of IG trade volume was printed in TRACE in the 5-minute interval around popular spotting times (i.e. 15:00, 15:30, 16:00).

[11] We do not drop trades that occur in common spotting times from the denominator when we compute the proportion of volumes that occurs in PT form. Therefore, our estimates of the proportion of TRACE volumes that occur in PT form are necessarily conservative.

Panel B in ***Table II*** shows how the portfolio trades identified by our algorithm have evolved over time. Portfolio trading activity has increased sharply, both in terms of the number of portfolio trades as well as the total dollar volume. We identified 1,950 unique PTs in 2018 and 2.5 times more in 2021 alone (4,914). Portfolio trading volume increased from $81 billion in 2018 to $311 billion in 2021. In percentage terms, the proportion of total trading volume that occurred in the form of PTs in 2021 was close to 7%, off a base of c.1% in 2018. This rapid growth demonstrates that the protocol has been quickly adopted by a large number of market participants, and justifies the requirement to add a PT flag to TRACE trades starting in May 2023.

Despite the large number of filters, we have put in place, the concern remains that due to its enormous size, TRACE contains many standard RFQ trades that are clustered by chance in ways that cause us to mischaracterize them as portfolio trades. However, were that to be the case, we should find a consistent flow of PTs in the TRACE data. Instead, the growth of the PT market as identified by our algorithm closely conforms to the growth in the volume of investor inquires, providing a validation check for our machine learning approach.

In ***Table III*** we check how well the algorithm identifies the portfolios in our Barclays inquiries database. For any given inquiry, the true positives rate is calculated as the number of line items the algorithm identified divided by the total number of line items in that inquiry. The false positives rate is calculated as the number of incorrectly identified line items divided by the total number of line items the algorithm found. With a median true positives rate of 97% and a false positives rate of 2.9% we are confident that the algorithm is successful at identifying actual portfolio trades in TRACE.

In ***Table IV***, we compute portfolio-level summary statistics along two dimensions: portfolio construction characteristics (Panel A, number of line items, volume, line item weights and sectors) and volume-weighted bond characteristics (Panel B, liquidity measured by *LCS*, maturity and bond age). We include statistics for both the full set of portfolios identified by our algorithm and the set of actual investor portfolio inquiries. The empirical distribution of the portfolio trades identified by the algorithm closely matches the distribution of investor inquiries in each of these key aspects, which suggests our algorithm is not mischaracterizing groups of RFQ trades as PTs. The average portfolio trade contains c.100 line items and $70 million worth of notional, approximately equally-split between the bonds in that portfolio. Portfolio trades are well-diversified and, on average span bonds from 12 different sectors. The average portfolio trade costs 0.83% to transact and is comprised of bonds with remaining maturity of about 10 years, issued 2.5 years ago.[12]

Finally, we perform our analysis of transaction costs using both the full dataset of PTs and the narrower set of PTs from the inquiry database, and find similar results, which is again supportive of our algorithm. We prefer the analysis using the full database of portfolio trades we construct using this algorithm because the larger sample size allows us to employ a more rigorous econometric specification and to explore in greater detail the cross-sectional heterogeneity in the data.

4. Crowd-sourcing Liquidity via Portfolio Trades

Next, we compare PTs identified by our algorithm to RFQs (defined as those trades not identified as PTs) on volume-weighted liquidity (LCS), maturity, and bond age (for RFQs see the last row *"TRACE ex PT"* of ***Table IV***). Along maturity and age, PTs are quite similar to RFQs. The main difference between them appears to be liquidity: the bonds traded in portfolio trades are substantially less liquid than the trades done using RFQs. The average LCS of the line items in PTs is 0.84% compared to 0.69% for RFQs (higher LCS implies

[12] As a further robustness check, in ***Figure 3*** we also overlay the percentage distribution of portfolio volumes by sector for the set of portfolio trades identified by the ML algorithm and the set of investor inquiries, and find no material differences in the sectoral distribution of volumes.

lower liquidity). Further, the distribution of portfolio LCS reveals that this is not driven by a few very illiquid portfolio trades. More than 50% of the portfolios, both by count and by volume, are less liquid than the average bonds traded in RFQs.

In ***Figure 4*** we expand on this result by showing how the aggregate distribution of IG portfolio volumes varies by LCS quintile. As a reference point, we also overlay on the same chart the distribution of volumes for bonds in the Bloomberg US IG Corporate Bond Index (BBG IG). The BBG IG Index contains about seven thousand bonds from a diverse set of issuers and measures the performance of the investment grade, fixed-rate, taxable US corporate bond market. The index is not skewed towards liquid bonds and is widely considered to be representative of the IG corporate bond market. ***Figure 4*** shows that PT volumes are shifted towards the less liquid quintiles. Compared to the BBG IG Index, there is 6% less portfolio volume in the first two most liquid quintiles, the majority of which then sits in the 4th LCS quintile.

Further, we find that investors construct these illiquid portfolios is a way that “crowd-sources” liquidity for the illiquid CUSIPs. To demonstrate this, for each portfolio we compute the percentage of portfolio volume contained in the two most liquid quintiles of LCS and in ***Figure 5*** plot the portfolio-level distribution of this percentage separately for liquid and illiquid portfolios. Liquid (Illiquid) PTs have lower (higher) trade volume-weighted LCS than the trade volume-weighted LCS of the bonds in the BBG IG Index. The boxplot shows that even amongst the illiquid PTs, very few PTs contain only illiquid bonds (the median percentage of liquid bonds in illiquid PTs is 37%). In other words, these portfolios appear to be designed such that the more liquid bonds cross-subsidize the less liquid bonds, resulting in an overall portfolio LCS that is closer to the index than if these illiquid bonds were traded individually.

We have also considered other possible portfolio construction strategies, including maturity, sector, and rating (***Table A3.1***). We compute either maturity, sector or rating-based Herfindahl scores (HHI), summing the squared percentages of trade volume for each individual portfolio, and compare these scores to the respective HHI score of the Bloomberg IG Corporate Bond Index. We classify portfolio trades into a maturity/sector/rating strategy if the HHI of the portfolio is at least 50% higher than the HHI of the Index. Among these other dimensions, a maturity-type strategy is the most common (35% of portfolios) and is typically focused on longer-dated bonds. However, 80% of these maturity-type portfolios can also be classified as illiquid. Even for portfolios tailored to a specific part of the market, liquidity remains a motivating factor.

5. Transaction Costs

5.1 Econometric Model and Identification

Our analysis of transaction costs uses data from 2021. Our motivation for using the most recent data only is that we want to study portfolio trading in a more mature stage of its development. Both the number and total volume of PTs in 2021 nearly equal the aggregate PT activity for all other years combined. We believe that the earlier data reflected instances when both investors and market-makers were getting familiar with the tool, and may not be representative of current PTs.

Following the literature (e.g. Bessembinder (2003); Collin-Dufresne, Junge, & Trolle (2020); Hagströmer (2021)), we measure transaction costs of trade *i* in bond *j* on day *t* by the effective half-spread (EHS):

$$EHS_{i,j,t} = D_{i,j,t}(P_{i,j,t} - M_{j,t})$$

where $D_{i,j,t}$ is an indicator variable that equals one for customer buy trades and negative one for customer sell trades, $P_{i,j,t}$ is the price at which the trade is executed and $M_{j,t}$ is the end-of-day mid-price as quoted by Bloomberg. *EHS* is an indication of how far traded prices are from the mid-price; values closer to the mid-price indicate lower transaction costs realized by investors.

To compare transaction costs for portfolio trades and non-portfolio trades, we estimate the following panel-data regression model at the transaction-level:

$$EHS_{i,j,t} = \beta_1 Portfolio\ Trade_{i,j,t} + \Gamma Z_{j,t} + \delta Block\ Trade_{i,j,t} + \lambda_j + \delta_t + \epsilon_{i,j,t} \quad \text{(Model 1)}$$

where $Portfolio\ Trade_{i,j,t}$ is a dummy variable equal to one when transaction i in bond j on date t is part of a portfolio trade. In our baseline analysis, $Portfolio\ Trade_{i,j,t}$ is defined on the sample of PTs identified by our clustering algorithm. $Z_{j,t}$ is a vector of time-varying bond characteristics (maturity and numeric rating (higher is worse)) and $Block\ Trade_{i,j,t}$ is a dummy variable which equals one if the notional traded in transaction i was greater than $5 million.[13] The specification also includes bond fixed effects (λ_j) and date fixed effects (δ_t). The inclusion of bond fixed effects controls for bond-level variables which do not vary over time, such as issue size, coupon, sector classification, whether a bond is callable, etc. The inclusion of date fixed effects controls for market-wide forces such as volatility, interest rates, the direction of the market (i.e., whether investors are net buyers or net sellers on a given day), which could systematically affect transaction costs on a given day. Finally, we cluster standard errors both at the bond and at the date level.

If portfolio trading is cost-effective, we would expect $\beta_1 < 0$: transaction costs are lower when a bond trades in portfolio trade compared to when the same bond trades in the standard RFQ protocol that day.[14] On any given trading date, there are three categories of bonds that

[13] We use a dummy instead of a continuous measure of quantity traded because we use the standard version of TRACE, where volumes for IG bonds greater than $5 million are capped.

[14] Using quotes data on S&P 500 stocks, Hagströmer (2021) shows that the level of the effective bid-ask spread measured relative to the mid-price could overstate the true bid-ask spread. The paper derives conditions under which EHS is not biased and proposes new estimators. While it is possible that Hagströmer's result also applies to the universe of corporate bonds, it does not compromise the validity of our results since our main interest is

appear in TRACE – (1) bonds which only trade in RFQ; (2) bonds which only trade in PTs; and (3) bonds which trade in both protocols (***Figure 6***). Identification of β_1 in ***Model 1*** comes from the sample of those bonds which have trades in both protocols. On average, approximately 21% of bond-date observations in our sample fall into this category, which is a meaningful portion and supports the empirical validity of our results. This statistic is fairly stable over time, with a minimum of 17% and a maximum of 25% (***Table V***).

5.2 Portfolio Trades Reduce Transaction Costs

We find that PTs are substantially more cost-effective than the standard RFQ protocol ($\beta_1 < 0$) (column (1) ***Table VI***). All else equal, the average transaction cost of a line item in a portfolio trade is 7.4 cents cheaper than the same trade in RFQ form. Given an average *EHS* of 16.5 cents for RFQs, the effect translates into a 44.6% reduction in transaction costs.

In column (2) of ***Table VII***, we re-estimate ***Model 1*** but define $Portfolio\ Trade_{i,j,t}$ using the sample of PT inquiries. Both the magnitude of the coefficients and their statistical significance remain unchanged. This is an important robustness check, which re-emphasizes that our algorithm identifies actual PTs.

We perform two other robustness checks, using both definitions of $Portfolio\ Trade_{i,j,t}$. First, we include bond-date fixed effects instead of two-way bond and date fixed effects and find very similar results (columns (1) and (3) in ***Table A3.2***). Second, we drop the fixed effects and saturate the model with a comprehensive set of bond-level and date-level controls (columns (2) and (4) in ***Table A3.2***). The magnitude of the estimates from these regressions tends to be slightly higher than our baseline. While it is relatively easy to control for time-varying features of bonds (e.g. maturity, rating etc.), modelling time-invariant features (which would be captured by the bond fixed effects) tends to be more difficult due to the complex

the estimate of the difference in execution costs between portfolio and non-portfolio trades, and not the level. By taking the difference in EHS, any bias will cancel out.

structure of these securities. For these reasons, we prefer the model with two-way fixed effects, which produces more conservative estimates.

5.3 The Cost Benefits Are Strongest for Illiquid Bonds

To examine how the benefit of PTs varies across bonds, we augment ***Model 1*** by adding a bond-level illiquidity term and its interaction with the $Portfolio\ Trade_{i,j,t}$ dummy:

$$EHS_{i,j,t} = \beta_1 Portfolio\ Trade_{i,j,t} + \beta_2 Illiq_{j,t} + \beta_3 Portfolio\ Trade_{i,j,t} \times Illiq_{j,t} + \Gamma Z_{j,t} + \delta Block\ Trade_{i,j,t} + \lambda_j + \delta_t + \epsilon_{i,j,t} \text{ (Model 2)}$$

where $Illiq_{j,t}$ is measured as one of: *LCS, TES, Bond Age, Price impact* or *Roll's measure* for bond *j* on date *t*. If portfolio trading reduces transaction cost to a greater extent for illiquid bonds, we would find both $\beta_1 < 0$ and $\beta_3 < 0$.

In this specification, we compare two differences: first, the difference in transaction costs when a bond is traded in a portfolio and when it is traded individually, and, second, the difference in transaction costs of an illiquid bond and a liquid bond. Hence, estimating β_3 relies on an additional source of variation compared to our baseline specification. We not only exploit variation in the transaction cost of a bond depending on the trade protocol, but also use cross-sectional variation in the liquidity profiles of the bonds we observe on any given day.[15] ***Figure 6*** demonstrates that the former holds. The distribution of portfolio volumes by LCS quintiles in ***Figure 4*** demonstrates the latter; portfolio trading is not exclusively confined to the very liquid or the very illiquid bonds only, but occurs across the entire spectrum of liquidity.

Table VII contains the results of ***Model 2***. Illiquid bonds incur higher transaction costs, irrespective of which protocol they are traded in ($\beta_2 > 0$). More importantly, the reduction in transaction costs for portfolio trades is higher for illiquid bonds ($\beta_1 < 0$ and $\beta_3 < 0$).

[15] This specification requires that we limit our analysis to the definition of $Portfolio\ Trade_{i,j,t}$ based on our algorithm because of the larger number of PTs identified.

Moreover, the magnitude of β_1 is substantially lower than the baseline estimate in column (1) of ***Table VI***, which shows that the benefits of portfolio trading are concentrated in illiquid bonds. We obtain statistically significant and qualitatively similar results regardless of the measure of illiquidity we employ.

To evaluate the economic magnitude of the effect, take for example two bonds: a liquid bond with low LCS = 0.5% and an illiquid bond with high LCS = 1%. Using the regression estimates in column (1) of ***Table VII***, we obtain that the reduction in transaction costs if traded in a portfolio for the liquid bond would be 6.89 cents (-4.61-4.55*0.5%); the reduction for the illiquid bond is nearly 50% greater at 9.16 cents (-4.61-4.55*1%).

Interestingly, ***Figure 6*** shows that in about 5% of our bond-date observations, we observe only PT volumes (i.e., there were no RFQ trades). These bonds are typically illiquid, with an average LCS score of 0.95% compared to a sample bond average of 0.69%. While estimating the causal benefit to transaction costs for trades which we only observe in portfolios is wrought with difficulties (since the counterfactual is "missing"), the fact that such trades exist in TRACE is significant by itself. It suggests that the benefits of portfolio trading for illiquid bonds extend beyond the reduction in transaction costs. There may be situations when there is no economically viable alternative, because either market-makers are not willing to provide immediacy or the transaction cost is prohibitively high, but investors nonetheless have a demand to trade. Portfolio trading could increase the chance that the trade will actually be executed, and thus improve the certainty of execution in addition to the transaction cost.

6. Relationship to the ETF Ecosystem

The extent to which portfolio trading reduces execution costs, particularly for illiquid bonds, raises the important question why this new trading protocol works so well. One hint about the possible source of the effectiveness of PTs comes from the high degree of overlap between the line items in PTs and the bonds held by ETFs. In ***Figure 7*** we plot the overlap with the largest IG ETF, the iShares iBoxx Investment Grade ETF (ticker: LQD). On average, 60% of the bonds in IG portfolio trades are owned by LQD. To put this in perspective, LQD owns about 30% of the bonds in the Bloomberg IG Corporate Bond Index. Over 90% of the portfolios in our sample have an overlap in excess of 30%, which shows that portfolio trades are significantly more concentrated in ETF bonds.

We identify two spill-overs from the ETF ecosystem that drive the reduced execution costs of PTs. First, market-makers leverage the ETF create and redeem process to offload the inventory of bonds which accumulates as a result of portfolio trading and/or to source bonds sold via PTs. Second, ETFs provide market-makers an intra-day pricing and hedging tool for transactions in diversified portfolios of corporate credit risk.

6.1 Relationship to the ETF C/R Process

One benefit of ETF ownership is that market-makers can use the ETF C/R process to either offload or source bonds. Bonds that are heavily owned by IG ETFs have a higher probability of being included in the daily ETF create or redeem baskets than bonds with low ETF ownership. Market-makers could deliver the bonds they bought from an investor to ETFs and create ETF shares, thus efficiently recycling the risk accumulated as a result of portfolio trading. Conversely, market-makers could redeem ETF shares and source bonds included in investor portfolios.

IG ETFs own a substantial number of illiquid bonds, despite the fact that they are typically benchmarked against liquid bond indices. This is because the liquidity of IG bonds declines quickly after issuance. We estimate that between 10%-15% of the bonds held by LQD in a given month belong the lowest quintile of liquidity, creating a natural demand for them. IG ETFs need these illiquid bonds in order to closely track the performance of their underlying benchmark. However, illiquid bonds might be difficult or very expensive to trade outside of portfolio trades, and hence, portfolio trades could provide a channel through which illiquid bonds can then be supplied to the ETFs.

To demonstrate the link to the ETF ecosystem, we explore variation in execution costs depending on the direction of the portfolio trade (i.e. customer sell or customer buy), and the composition of the ETF C/R process on the day when the portfolio trade was executed. If the ETF C/R process plays a role in the transaction costs of portfolio trades, then we would expect that market-makers would put better prices on illiquid bonds that they believe can be packaged with other debt and offloaded in a timely manner via the C/R process.

Focusing on bonds owned by LQD, we define a trade in an individual bond that an investor sells to a market-marker as "right way" for the LQD when the bond is part of the create basket that day, which would allow market-makers to deliver it to the LQD and create shares. Conversely, an investor buy trade is "right way" when the bond is part of the redeem basket that day, in which case the market-makers could redeem LQD shares and source the bond. We estimate the following transaction level regression:

$$EHS_{i,j,t} = \beta_1 Portfolio\ Trade_{i,j,t} + \beta_2 Right\ way^{LQD}_{i,j,t} + \beta_3 Portfolio\ Trade_{i,j,t} \times Right\ way^{LQD}_{i,j,t} + \Gamma Z_{j,t} + \delta Block\ Trade_{i,j,t} + \lambda_j + \delta_t + \epsilon_{i,j,t} \quad \text{(Model 3)}$$

In this model, the coefficient β_2 measures how being "right way" for LQD alters execution costs on all trades and β_3 gives the incremental impact of being right way for LQD for a trade that is executed in a PT. Following Koont et al. (2022), in order to reduce noise in the data, we do not impute C/R baskets on days when ETF portfolio changes are very small.[16]

Table VIII presents the results of ***Model 3.*** Regardless of which protocol they are executed in, trades which are "right way" for LQD achieve on average better execution than

[16] We restrict our sample to days when the daily percentage change in the number of ETF shares was below the 25th percentile (i.e. significant redeems) or above the 75th percentile (significant creates).

trades that are "wrong way" ($\beta_2 < 0$ in column (1)). However, the effect is 60% stronger for "right way" trades executed in the PT protocol ($\beta_3 < 0$ in column (1)).

We next split the universe of LQD bonds in three terciles (liquid, medium liquid and illiquid), and re-evaluate the model. The magnitude of the "right way" effect (β_2) is the lowest for the liquid tercile (column (2) in ***Table VIII***): these bonds are relatively easy to trade and so are less reliant on ETF activity. Further, the incremental benefit of being right way in a PT versus an RFQ (β_3) is low (at only 2 bps of *EHS*) and statistically insignificant. In contrast, the benefit of being right way is strongest for the least liquid tercile (column (4) in ***Table VIII***). These bonds are generally difficult to trade and thus benefit substantially from alignment with ETF trading activity. More importantly, the incremental benefit of being right way in a PT is high (a 10 bps reduction in *EHS*) and statistically significant.

Recall from the prior section that the benefits of PTs are most concentrated in illiquid bonds. It is precisely for these illiquid bonds that the effectiveness of PTs is the most sensitive to alignment with the direction of ETF C/R activity: being right way results in substantially lower execution costs. Combining these two pieces of evidence, we conclude that PTs benefit from being aligned to C/R activity to a far greater extent than RFQs.

As a robustness check, we have estimated another version of ***Model 3***, where we use the general direction of the ETF C/R process (i.e., net create or net redeem) rather than a bond-level measure. We classify each day in our sample as a net create or net redeem day depending on whether LQD experienced net inflows or net outflows, omitting by construction days with zero flows. We define any investor sell (buy) trade as "right way" when the ETF is in create (redeem) mode. We estimate the model using both TRACE PTs (***Table A3.3***) and client inquiries[17] (***Table A3.4***) and find that our results and conclusions

[17] Using the inquiries database, the magnitudes of the effects are similar, but the coefficients are not statistically significant. Nonetheless, this is an encouraging result and one that is not surprising given the demanding

remain qualitatively unchanged: portfolio trades benefit more from being aligned with the ETF C/R process than do comparable RFQ trades.

6.2 Relationship to the ETF Prices

In principle, both PTs and RFQs could benefit from using the ETF C/R process to offload risk or source bonds. The negative and statistically significant coefficient on $Right\ way^{\mathrm{LQD}}_{\mathrm{i,j,t}}$ in ***Table VIII*** certainly corroborates this intuition: trades that are "right way" for the ETF generally incur lower transaction costs, particularly for illiquid bonds. Yet the importance of being "right way" is stronger for PTs. Therefore, an open question remains: why does packaging bonds into portfolios reduce transaction costs?

We believe the answer is linked to another feature of ETFs: they trade actively in the secondary market. Intra-day trading provides both price transparency and hedging tools that are far more applicable to portfolios of corporate bonds than to individual bonds. Where the intra-day price of an ETF is clearly relevant for the price of a diversified portfolio of corporate bonds, particularly one with a high degree of overlap with an ETF, the pricing of a single security will reflect mostly idiosyncratic risks and security-specific supply and demand. Similarly, a diversified basket of corporate bonds can be hedged using an ETF, whereas hedging a single name position with an offsetting position in an ETF incurs substantial basis risk.

If we are correct that the degree of transparency and hedging flexibility afforded to PTs by ETFs allows market-makers to reduce PT transaction costs, then the execution prices of PTs should mirror more closely the intra-day variation in the price of ETFs than do the execution prices of RFQs. To test this hypothesis, we turn to the secondary dataset of investor

---

econometric specification. The model requires that we observe the same bond in at least two portfolio trades per day: one that is "right way" for LQD and another that is "wrong way" for LQD. In the full dataset computed using our algorithm, approximately 7% of the bond-dates fulfil this criterion, whereas this applies to less than 2% of the bond-dates in the investor inquiries database (further motivating our use of the full database).

portfolio inquiries. This sample is smaller but contains the exact time stamp of the inquiry, as well as the Bloomberg quoted bid/offer price for each line item at that time. We combine these bond quotes with the corresponding quote on LQD at that precise time and compute ETF deviations from NAV. For an illustrative example of the data, see Panel B of ***Table I***.

This data allows us to assess the intra-day execution of PTs, but not of RFQs. In order to compare the intra-day pricing of the two protocols, we do the following exercise. For each bond in a portfolio trade, we define the implied RFQ price as the average of all RFQ transactions (weighted by notional) recorded in TRACE for that bond, in the same direction as the PT, that were executed within four hours of the PT.[18]

For each line item, we now have the realized PT price, an implied RFQ price, and the Bloomberg quoted price at the time of the inquiry. We then convert both the PT price and the implied RFQ price into a measure of deviation from the Bloomberg quote. Essentially, this is a real-time analogue of the *EHS* based on end-of-day quotes that we used earlier. More specifically, for each line item i in a given portfolio p, we measure the transactions costs, computed in basis points (bps) as:

$$PT\ Dev\ from\ Quote_{i,p} = \begin{array}{l} \dfrac{P_{i,p} - Bid_i}{Bid_i}\ for\ investor\ sell\ trades, bps \\ \dfrac{Offer_i - P_{i,p}}{Offer_i}\ for\ investor\ buy\ trades, bps \end{array}$$

where $P_{i,p}$ is the traded price recorded in TRACE and $Bid_i/Offer_i$ is the Bloomberg quote at the exact time of the day when the portfolio was received. Similarly, we compute the Implied RFQ transaction cost as:

[18] Choosing a shorter window improves the estimated price of coincident RFQ trades but limits the sample of PT line items with a matched RFQ trade.

$$Implied\ RFQ\ Dev\ from\ Quote_{i,p} = \begin{cases} \dfrac{Implied\ P_{i,p} - Bid_i}{Bid_i} & for\ investor\ sell\ trades, bps \\ \dfrac{Offer_i - Implied\ P_{i,p}}{Offer_i} & for\ investor\ buy\ trades, bps \end{cases}$$

where $Implied\ P_{i,p}$ is the implied RFQ price for that line item, in place of $P_{i,p}$. Finally, to arrive at a portfolio-level measure of transaction costs, $PT\ Dev\ from\ Quote_p$ or $Implied\ RFQ\ Dev\ from\ Quote_p$, we average across the transaction costs of all line items in a given portfolio trade, using the portfolio trade notionals as weights. Higher $Dev\ from\ Quote_p$ signify lower transaction costs – the higher prices are above the bid quote when investors buy and the lower prices are below the offer quote when investors sell, the better is the execution outcome for investors.

We also build the corresponding measure of ETF deviations from NAV, conditioned on the direction of the portfolio trade, using the quoted ETF price at the exact time stamp of the portfolio inquiry:

$$ETF\ Dev\ from\ NAV_p = \begin{cases} \dfrac{ETF\ Price - NAV}{NAV} & for\ investor\ sell\ trades, bps \\ \dfrac{NAV - ETF\ Price}{NAV} & for\ investor\ buy\ trades, bps \end{cases}$$

We then estimate the following portfolio-level regression:

$$Dev\ from\ Quote_p = \alpha + \beta_1 ETF\ Dev\ from\ NAV_p + \epsilon_p \text{(Model 4)}$$

separately for each version of $Dev\ from\ Quote_p$ and report the results in ***Table IX***. First, as expected, we find that PT execution costs are strongly positively correlated to ETF deviations from NAV. The magnitude of the effect is economically meaningful. When investors sell bonds via PTs at a time when the ETF is trading one basis point above NAV, the PTs are executed on average 0.58 basis points above the Bloomberg bid price (column (1) in ***Table IX***). Conversely, when investors sell and the ETF trades one basis point below NAV, PTs execute 0.58 basis points below the bid price. These results are consistent with our previous conclusion that transaction costs are lower when the PTs are aligned with the direction of ETF flows. The same effect holds when investors buy bonds via PTs when the ETF trades below NAV.[19]

In column (2), we augment ***Model 4*** by including several portfolio-specific characteristics: liquidity, tail liquidity (difference between the liquidity of ETF and non-ETF bonds), volume and number of line items. Interestingly, adding these variables improves only marginally the R-squared of the regression, leading us to conclude that ETF deviations from NAV is by far the most important factor driving PT prices.

We now test if the implied RFQ prices have a different sensitivity to ETF deviations from NAV. Since computing this is only meaningful for those portfolios where we see all or (nearly all) line items in both protocols on the same day (about 60% of the PTs), we first verify that the sensitivity to ETF deviations is not systematically different for those PTs which have an RFQ equivalent compared to those that do not (column (3) in ***Table IX***).

Then, we estimate ***Model 4*** using implied RFQ prices and report the results in columns (4) and (5). The sensitivity to ETF deviations from NAV is one fourth as large (β_1=0.15) and the coefficient is not statistically significant. Further, although the total R-squared is similar across the two specifications, the contribution of ETF and portfolio-specific factors is materially different. While movements in the ETF price explain the majority of variation in PT prices, the factors related to bond characteristics drive RFQ execution outcomes.

Our results show that for the same bonds traded on the same day, PT and RFQ protocols are priced in very different ways, which helps answer the two questions we posited in this paper: (1) Why trade in a portfolio and not individually, and (2), Why is portfolio trading so

---

[19] Because of our limited sample size, we estimate the regressions by pulling all trades together. However, we verify in robustness checks that the direction and magnitude of the estimate are comparable for investor sells and for investor buys.

cost effective for illiquid bonds? ETFs provide benefits of price transparency, price discovery and risk hedging, which only apply when the bonds are traded in portfolio form. The relationship to the ETF ecosystem is especially important for older, less liquid bonds. These bonds tend to trade very infrequently, as a result of which their price marks could be stale. Exploiting the high overlap between the line items in a portfolio trade and the ETFs improves the price discovery process for these illiquid securities.

6.4 Alternative Explanations

The main focus of our paper is to the estimate the difference in execution costs between PT and RFQ trades, and not the level. In fact, the time-series correlation between the execution costs of these two protocols is 0.80, which suggests the presence of common factors driving the level of costs for both protocols. Market-makers' bid-offer spreads reflect a number of components, which have been well documented in the literature: inventory and hedging costs (e.g. Goldstein and Hotchkiss (2020)), price transparency (e.g. Edwards, Harris, & Piwowar (2007)), and volatility among others. However, the cost benefit of portfolio trading cannot be attributed to these common factors, since by taking the difference in *EHS*, their effect will cancel out. Based on the evidence we provide, we believe that ETFs provide the most likely explanation why portfolio trades work so well. Nonetheless, we have also considered two alternative explanations– competition for market share and clients "swapping" portfolios. We find strong evidence against either of these theories.

Market-makers might use portfolio trades to gain market share, and the competition to win these trades could drive the bid-offer they charge. According to this narrative, the reduction in execution costs could either reflect an increased motivation to win the trades, or it could be linked to some information that market-makers obtain through executing these trades that they would not obtain by unsuccessfully bidding/offering on the same portfolio. To test this theory, in ***Figure 8*** we plot the time series of the difference between the average

EHS of transactions executed via the PT and RFQ protocol over the period Jan 1st 2020 – December 31st 2021. We also show PT volumes as a percentage of total TRACE IG volumes. The reduction in *EHS* associated with PTs remained stable over a period when the use of portfolio trading rose dramatically. This is the opposite of what we would expect if marker-markers were simply "buying" market share, as they would have increased the benefit of portfolio trading to entice more participation.

Another explanation could be that investors use PTs to "swap" entire portfolios with other investors, who want to trade the same bonds but in the opposite direction. In this scenario, a market-maker would act as an agent, lining up both sides of the PT and charging a considerably lower bid-offer than if the market-maker had acted as a principal and priced the bonds out of inventory. If this were common, we would find examples of offsetting PT trades in our database. However, we find that this occurs in less than 0.5% of the PTs in our sample (***Table X***), which speaks strongly against the theory.

7. Discussion

In this article we introduce the concept of portfolio trading, the latest innovation in the corporate bond market, which involves trading a basket of bonds as a single piece of risk, and transacting the entire basket with one single market-maker. Using a proprietary dataset of portfolio inquiries, we develop an algorithm to identify corporate bond portfolio trades in TRACE. We show that investors typically use this new trade protocol to transact in illiquid bonds and that portfolio trades reduce transaction costs by more than 40% compared to trades using the standard RFQ protocol. We also demonstrate that spill overs from the ETF ecosystem allow market-makers to both offload the inventory of bonds which accumulates as a result of portfolio trading and provide a transparent intra-day tool to price and hedge PTs.

Our work opens a broad set of avenues for future research. It will be interesting to investigate if and how our results translate to the HY market. HY ETFs tend to be even more focused on the liquid spectrum of bonds that IG funds, implying that HY ETF managers should theoretically have a lower demand for illiquid HY bonds than IG ETF managers. A promising research question one could ask is whether the HY ETF C/R mechanism also helps to reduce execution costs for illiquid HY bonds.

Jiang, Li and Wang (2021) show that when faced with significant redemptions investors typically follow a pecking order of liquidity, selling the most liquid assets first. Meli and Todorova (2022) show that institutional investors use corporate bond ETFs to manage flows-driven liquidity, thus increasingly substituting bond trade volumes with ETFs. In this article, we provide evidence that compared to the standard RFQ volumes, PTs facilitate much more efficient trading in less liquid bonds. Together, these findings suggest greater capacity to manage trading needs at lower cost. The natural question to ask is how the liquidity risk premium has responded to the change in the demand for liquidity as investors have adopted both new products and new ways to manage liquidity and transact in illiquid bonds.

Another open question is if the rise in PT volumes itself generates spill-overs. The inclusion of less liquid CUSIPs in PTs may give market-makers more comfort providing liquidity in that part of the market, even away from PTs, thereby causing liquidity to snowball. Finally, we are interested in how transaction costs vary with market conditions. Our analysis offers some preliminary guidance in that direction, but it would be helpful to explore this topic in further detail. For example, periods when trade volumes are singularly one-sided (investors are either heavily net buying or heavily net selling) typically coincide with periods when the ETF C/R mechanism is also one-sided, but in a way that most of the trades are “wrong way” for the ETFs. This could potentially limit the ability to offload risk via the C/R mechanism and, thus, also could limit the cost effectiveness of portfolio trades.

On a related vein, periods of market distress also typically correlate heavily with periods when the volatility of the ETF bid/offer sharply increases, translating into higher hedging costs of PTs. Market-makers in their turn are likely to pass on these costs to investors, which could result in higher transaction costs of PTs.

List of Figures

Figure 1: Client Inquiries

The figure shows the growth in the number and $ volume of investor portfolio inquiries received by Barclays trading desk.

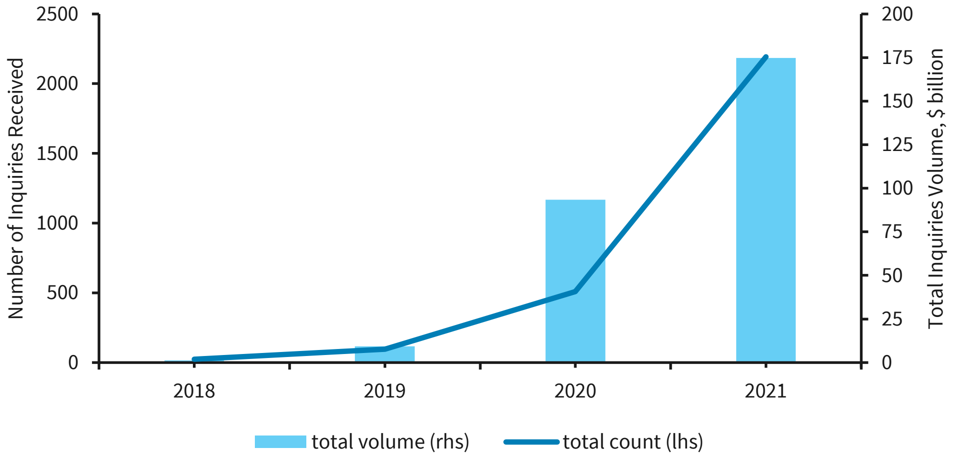


Figure 2: Flowchart of the Methodology Process

The figure shows the steps we undertook to construct the dataset of TRACE portfolio trades.

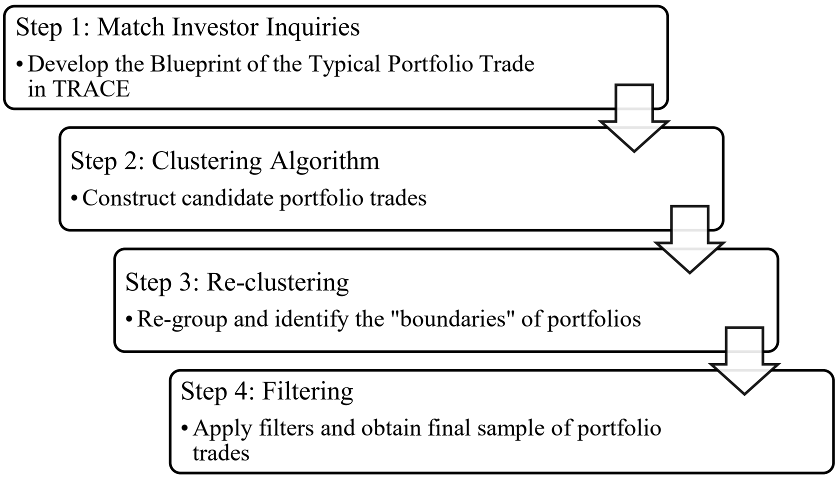


Figure 3: Algorithm Validation – Distribution of Portfolio Trade Volumes by Sector

The figure overlays the distribution of volumes by sector for the TRACE portfolios identified by the ML algorithm and the investor inquiries. Data based on portfolio trades executed during the period 1st January 2021 – 31st December 2021.

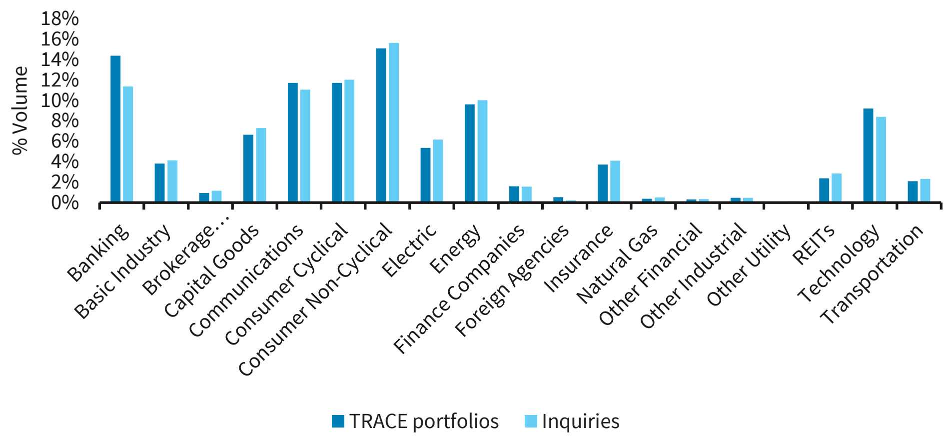


Figure 4: Distribution of Portfolio Volume by Liquidity Quintile

The figure shows the distribution of total IG portfolio trade volume by LCS quintile, where Q1 comprises the most liquid bonds and Q5 comprises the least liquid bond. Data based on portfolio trades identified by our ML algorithm executed during the period 1st January 2021 – 31st December 2021.

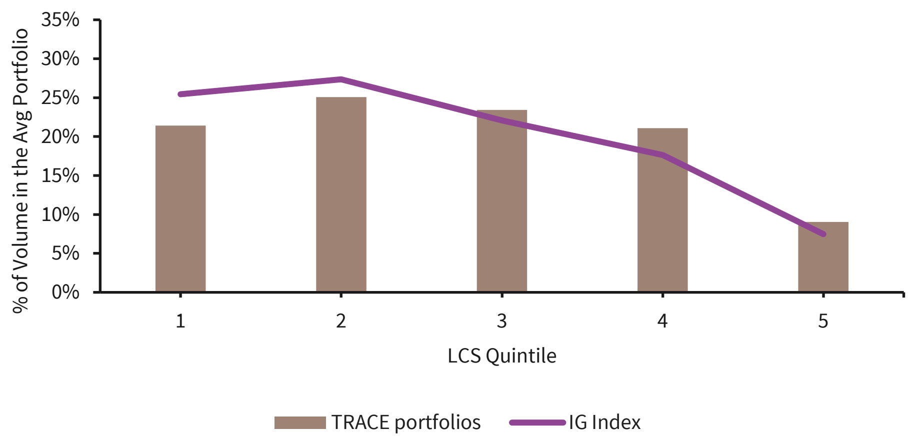


Figure 5: Mixing Liquid and Illiquid Bonds

The boxplot shows the distribution of the percentage of liquid volume (sum of trade volumes in the first two most liquid LCS quintiles) for Liquid and Illiquid portfolio trades. Liquid (Illiquid) PTs have lower (higher) trade volume-weighted LCS than the trade volume-weighted LCS of the bonds belonging to the Bloomberg IG Corporate Bond Index. Lower (higher) LCS is better (worse). Each box gives the 25th, median (red line) and 75th percentile of the distribution. Data based on portfolio trades identified by our ML algorithm executed during the period 1st January 2021 – 31st December 2021.

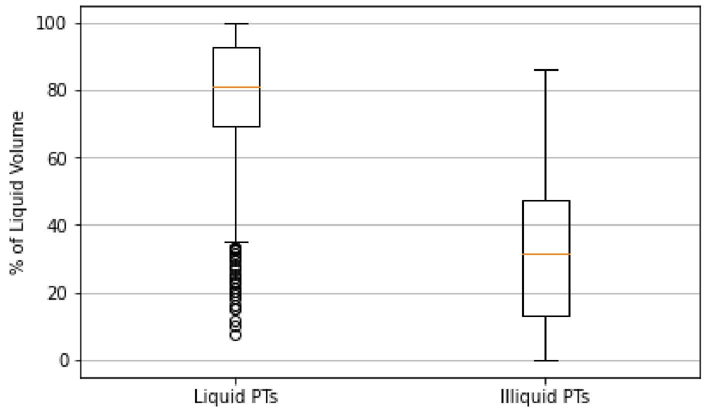


Figure 6: Identification Strategy

The figure shows the distribution of bond-date fraction of portfolio trades, expressed as a percentage of total count of trades on that day. Data based on portfolio trades identified by our ML algorithm executed during the period 1st January 2021 – 31st December 2021.

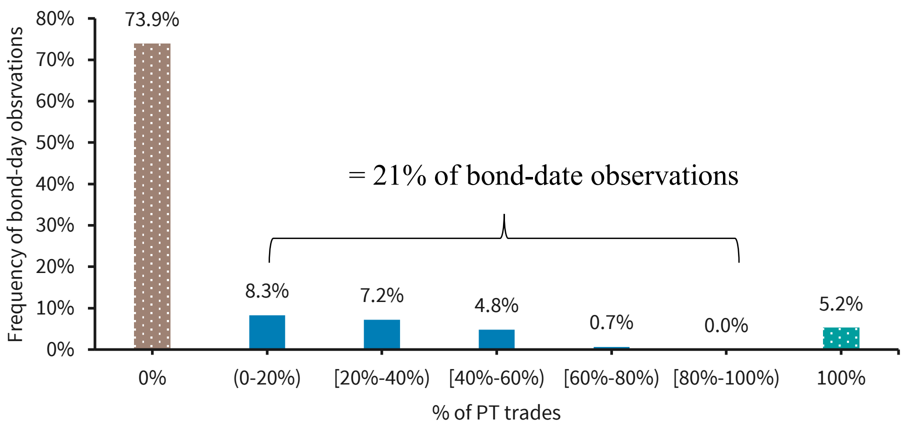


Figure 7: **Overlap Between Portfolio Trades and ETFs**

The figure shows the overlap between the line items of IG portfolio trades and the monthly holdings of LQD. An overlap of 0 means that *none* of the bonds in a given portfolio trade are held by LQD in that month; conversely, an overlap of 1 means that all of the bonds in the portfolio are held by LQD in that month. Data based on portfolio trades identified by our ML algorithm executed during the period 1st January 2021 – 31st December 2021.

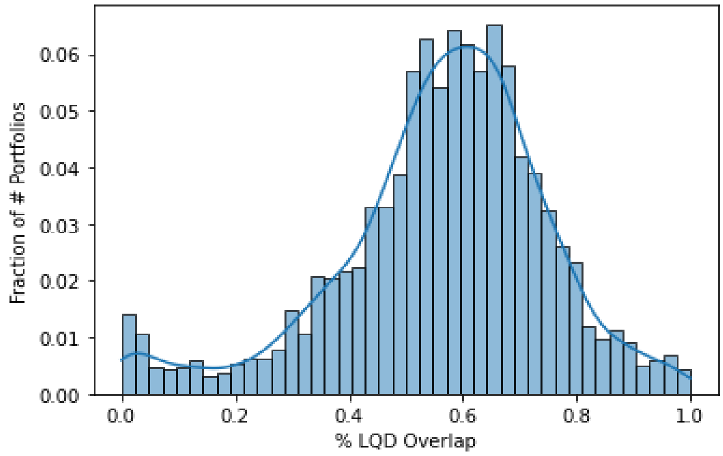
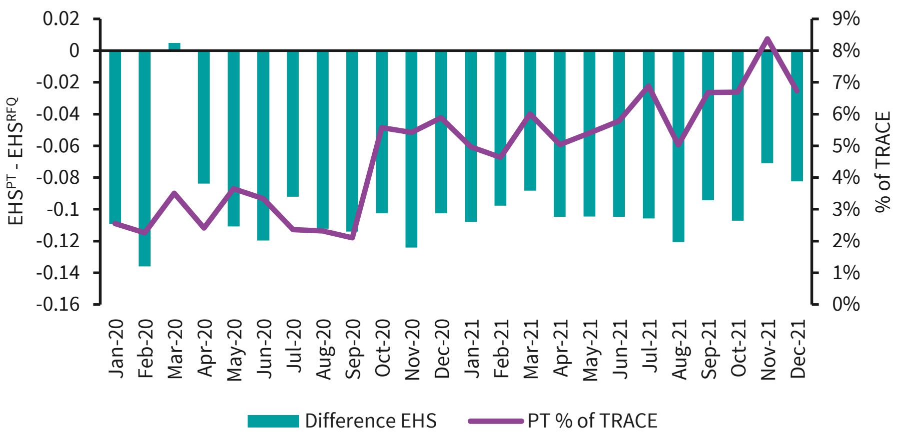


Figure 8: **Portfolio Trade Execution Costs Over Time**

The figure shows the difference between the average EHS of transactions executed in the PT and RFQ protocol (LHS) and the PT volumes as a percentage of TRACE (RHS). Smaller values of the difference indicate lower transaction costs of the PT protocol compared to the RFQ protocol. For more details on the definition of our TRACE universe refer to Section 2.3. Data are based on portfolio trades identified by our ML algorithm executed during the period 1st January 2020 – 31st December 2021.



List of Tables

Table I: A Portfolio Trade Examples

The table gives two examples of portfolio inquiries. The first inquiry comes from our primary dataset has 145 line items (Panel A); the second inquiry comes from our secondary (intra-day) dataset and has 250 line items (Panel B). Note that since data are proprietary, all values displayed in the table are for illustrative purposes only and do not represent actual inquiries.

Panel A: Primary Dataset

Date	PT ID	PT ID_Cusip	Cusip	Quantity	Direction
2021-01-05	123	123_1	05971KAE9	$250,000	Client Buys
2021-01-05	123	123_2	03835VAG1	$500,000	Client Buys
2021-01-05	123	123_3	037833CJ7	$750,000	Client Buys
2021-01-05	…	…	172967LD1	…	…
2021-01-05	123	123_144	29444UBE5	$300,000	Client Buys
2021-01-05	123	124_145	404119BN8	$500,000	Client Buys

Panel B: Secondary Dataset (Intra-day Data)

Time stamp	PT ID	PT ID_Cusip	Cusip	Quantity	Direction	BVAL Bid	BVAL Offer	ETF Dev from NAV
2022-09-12 11:32:42 EST	348	348_1	874054AH2	$700,000	Client Sells	$92.41	$92.71	7.62 bps
2022-09-12 11:32:42 EST	348	348_2	8426EPAF5	$250,000	Client Sells	$100.63	$100.72	7.62 bps
2022-09-12 11:32:42 EST	348	348_3	855244AU3	$100,000	Client Sells	$87.97	$88.27	7.62 bps
…	…	…	86765BAV1	…	…	$87.23	$87.61	7.62 bps
2022-09-12 11:32:42 EST	348	348_249	87264AAX3	$550,000	Client Sells	$87.52	$88.06	7.62 bps
2022-09-12 11:32:42 EST	348	348_250	87264ACB9	$600,000	Client Sells	$83.13	$83.22	7.62 bps

Table II: The Portfolio Trades Database

The table presents summary statistics of the portfolio trades database constructed using our ML algorithm. The estimate of the TRACE market excludes non-index corporate bonds, but includes volumes at common spotting times. For more details on the bond sample and discussion around spotting times, refer to Section 2 and Section 3.

	# Bond-PT Obs.	**# of PTs**	**$ Volume (bln)**	**% of TRACE**
		Panel A: Aggregate		
2018-2021	998,975	12,107	696	3.47
		Panel B: Time Series		
2018	107,541	1,950	81	1.14
2019	175,224	2,265	127	1.68
2020	245,774	2,978	177	3.09
2021	470,436	4,914	311	6.89

Table III: Algorithm Validation – True Positives and False Positives

The table shows how well the clustering algorithm is capable of identifying the "true" portfolios in the investor inquiries database. For any given inquiry, the true positives rate is calculated as the number of line items the algorithm identified divided by the total number of line items in that inquiry. The false positives rate is defined as the number of incorrectly identified line items divided by the total number of line items the algorithm found.

	Portfolio-level:	
	True positive rate	False Positive rate
Mean	85%	15%
Median	100%	3%

Table IV: Algorithm Validation – Empirical Distribution

The table compares the empirical distributions of the investor inquiries (INQ) and the portfolio trades (PT) identified using our ML algorithm along two dimensions: characteristics of the portfolio (Panel A) and characteristics of the bonds in the portfolio (Panel B). Portfolio LCS, Maturity and Age (time since issuance) are computed as a weighted average where the weights are given by the notionals of the line items in that portfolio. Our sample of inquiries comprises those inquiries which we could successfully match in full to the TRACE database. The last low, *TRACE ex PT,* reports the volume-weighted LCS, maturity and bond age for all non-portfolio trades in TRACE.

	Panel A: Portfolio Characteristics								**Panel B: Bond Characteristics**					
	# Line Items		Volume ($ mn)		Line Item Wgt (%)		# of Sectors		LCS (%)		Maturity (years)		Bond Age (years)	
	INQ	PT	INQ	PT	INQ	PT	INQ	PT	INQ	PT	INQ	PT	INQ	PT
Mean	93	97	76.3	68	2.16	2.04	11	12	0.83	0.84	9.44	10.22	2.53	2.62
Std	114.9	115.67	118.6	109	1.73	2.93	3.45	3.05	0.29	0.45	6.01	5.94	1.32	1.38
P25	27	37	14.1	21	0.83	0.97	9	10	0.68	0.59	6.15	6.21	1.66	1.71
Median	51	60	36.2	34.4	1.75	1.72	11	12	0.81	0.75	7.1	8.13	2.25	2.42
P75	109	105	89.9	69	3.12	2.78	14	14	0.92	0.97	10.66	12.72	3.12	3.17
TRACE ex PT	*NA*								0.69		10.7		2.44	

Table V: Identification Strategy – Time-series and Cross-sectional Distribution

The table shows the daily time-series distribution of the percentage of bonds that traded in both RFQ and portfolio trade protocols over the period Jan 1st 2021-December 2021 (column (1)) and for those bonds that traded in both, the average number of portfolio trades they were included in per day (column (2)).

	% of Bonds with both RFQ and PT Trades	**# of PTs for bonds with at least 1 PT**
Mean	21.0%	1.15
Std	6.7%	0.25
P25	16.5%	1.0
Median	20.8%	1.1
P75	41.6%	1.2
Observations	250 days	10,622 bonds

Table VI: Transaction Costs of Portfolio Trades

$$\boldsymbol{EHS_{i,j,t} = \beta_1 Portfolio\ Trade_{i,j,t} + \Gamma Z_{j,t} + \delta Block\ Trade_{i,j,t} + \lambda_j + \delta_t + \epsilon_{i,j,t}}$$

The table reports transaction-level regressions of effective half spread on a portfolio trade dummy ($Portfolio\ Trade_{i,j,t}$) and a set of controls. Regressions control for bond-level maturity and numeric rating (higher values are worse) (collected in vector $Z_{j,t}$) and include a transaction-level dummy variable equal to 1 for all trades larger than \$5 million ($Block\ Trade_{i,j,t}$). All variables are winsorized at the 1% level. All regressions include bond and date fixed effects. T-stats in parentheses. Standard errors are clustered at the bond and date (day) level. Significance at the 1 %, 5 % and 10 % statistical level is denoted by ***, **, and * respectively.

	Effective Half Spread (EHS)	
	(1) TRACE Portfolios	(2) Investor Inquiries
Portfolio Trade	-7.37*** (-22.71)	-7.63*** (-5.24)
Maturity	0.62*** (7.58)	0.59*** (715)
Numeric rating	0.55*** (2.79)	0.57** (2.17)
Block Trade	-4.36*** (-17.65)	-3.91*** (-16.10)
Mean EHS (Portfolio Trade = 0)	16.5	15.8
% Improvement	44.6%	48.3%
Bond FE	YES	YES
Date FE	YES	YES
Bond-trade Observations	4,467,324	
Sample Period	Jan 1st 2021 – December 31st 2021	

Table VII: Transaction Costs of Portfolio Trades By Liquidity Profile

$$\boldsymbol{EHS_{i,j,t} = \beta_1 Port\,Trade_{i,j,t} + \beta_2 Illiq_{j,t} + \beta_3 Port\,Trade_{i,j,t} \times Illiq_{j,t} + \Gamma Z_{j,t} + \delta Block\,Trade_{i,j,t} + \lambda_j + \delta_t + \epsilon_{i,j,t}}$$

The table reports transaction-level regressions of effective half spread on a portfolio trade dummy ($Port\ Trade_{i,j,t}$), bond-level illiquidity ($Illiq_{j,t}$) and their interactions. Regressions control for bond-level maturity and numeric rating (higher values are worse) (collected in vector $Z_{j,t}$) and include a transaction-level dummy variable equal to 1 for all trades larger than $5 million ($Block\ Trade_{i,j,t}$). All variables are winsorized at the 1% level. All regressions include bond and date fixed effects. T-stats in parentheses. Standard errors are clustered at the bond and date (day) level. Significance at the 1 %, 5 % and 10 % statistical level is denoted by ***, **, and * respectively.

	Effective Half Spread (EHS)				
	(1) LCS	(2) TES	(3) Bond Age	(4) Price Impact	(5) Roll
Portfolio Trade	-4.61***	-5.57***	-5.68***	-5.63***	-2.04**
	(-16.86)	(-12.03)	(-6.68)	(-14.91)	(-1.96)
Illiquidity	3.32***	0.33***	1.08	0.19***	4.11***
	(4.31)	(7.98)	(1.55)	(15.54)	(4.39)
Portfolio Trade ×Illiquidity	-4.55***	-0.51***	-0.51***	-0.36***	-4.65***
	(-6.39)	(-6.73)	(-3.51)	(-12.68)	(-6.19)
Maturity	0.46**	0.66***	0.50***	0.39**	0.55***
	(2.47)	(3.97)	(3.11)	(2.08)	(3.27)
Numeric rating	0.13	0.05	0.49*	0.59*	0.40
	(0.81)	(0.96)	(1.76)	(1.89)	(1.27)
Block Trade	-4.33***	-3.92***	-4.36***	-4.31***	-4.36***
	(-13.72)	(-12.00)	(-14.34)	(-14.06)	(-14.41)
Bond FE	YES	YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES
Bond-trade Observations	4,467,324	4,403,651	4,863,671	4,745,469	4,863,671
Sample Period	Jan 1st 2021 – December 31st 2021				

Table VIII: Relationship to the ETF Ecosystem – "Right way" for the ETF C/R

$$\boldsymbol{EHS_{i,j,t} = \beta_1 Port\,Trade_{i,j,t} + \beta_2 Right\,way^{LQD}_{i,j,t} + \beta_3 Port\,Trade_{i,j,t} \times Right\,way^{LQD}_{i,j,t} + \Gamma Z_{j,t} + \delta Block\,Trade_{i,j,t} + \lambda_j + \delta_t + \epsilon_{i,j,t}}$$

The table reports transaction-level regressions of effective half spread on a portfolio trade dummy ($Port\ Trade_{i,j,t}$), a trade-level dummy variable ($Right\ way_{i,j,t}$) and their interactions for bonds owned by the largest IG ETF (ticker LQD). $Right\ way_{i,j,t}$ equal 1 for all customer sell (buy) trades that also belong to the imputed LQD create (redeem) basket and zero otherwise. Since our methodology for imputing create (redeem) baskets computes average baskets, to minimize noise we focus only on days when LQD is heavily creating (redeeming) shares (i.e. daily percentage change in ETF shares is below the 25th percentile or above the 75th percentile). We split our sample into three sub-samples using terciles of the bond-level LCS distribution: liquid, medium and illiquid. Regressions control for bond-level maturity and numeric rating (higher values are worse) (collected in vector $Z_{j,t}$) and include a transaction-level dummy variable equal to 1 for all trades larger than $5 million ($Block\ Trade_{i,j,t}$). All variables are winsorized at the 1% level. All regressions include bond and date fixed effects. T-stats in parentheses. Standard errors are clustered at the bond and date (day) level. Significance at the 1 %, 5 % and 10 % statistical level is denoted by ***, **, and * respectively.

	Effective Half Spread (EHS)			
	(1) All LQD Bonds	(2) Liquid	(3) Medium	(4) Illiquid
Portfolio Trade	-6.55*** (-4.17)	-4.38*** (-4.76)	-5.75*** (-3.28)	-10.6*** (-3.87)
Right way	-10.10** (-2.33)	-6.22*** (-2.76)	-8.65** (-2.15)	-14.42* (-1.83)
Portfolio Trade × Right way	-6.64* (-1.78)	-2.63 (-1.32)	-5.57 (-1.43)	-9.56* (-1.73)
Maturity	0.22 (0.43)	-0.29 (-0.57)	1.42*** (2.66)	0.52 (0.35)
Numeric rating	-0.73 (-0.72)	-0.13 (-0.15)	-2.20 (-1.50)	-0.34 (-0.11)
Block Trade	-3.79*** (-4.51)	1.96*** (-2.79)	-2.80*** (-2.78)	-5.14*** (-3.06)
Bond FE	YES	YES	YES	YES
Date FE	YES	YES	YES	YES
Bond-trade Observations	185,784	87,784	64,207	49,969
Sample	Bonds included in LQD Jan 1st 2021 – December 31st 2021			

Table IX: Relationship to the ETF Ecosystem – ETF Prices

$$\boldsymbol{PT\ Dev\ from\ Quote_p = \alpha + \beta_1 ETF\ Dev\ from\ NAV_p + \beta_2 PT\ LCS_p + \beta_3 Tail\ Liquidity_p + \beta_4 Volume_p + \beta_5 Volume_p + \epsilon_p}$$

The table reports portfolio-level regressions of intra-day deviations from Bloomberg bid/offer quotes on intra-day ETF deviations from NAV, portfolio LCS (computed from bond-level LCS weighted by notional), tail liquidity (difference between the notional-weighted LCS of bonds not held by LQD and those held by LQD), portfolio volume and number of line items. T-stats in parentheses. Significance at the 1 %, 5 % and 10 % statistical level is denoted by ***, **, and * respectively.

	Deviations from Bid/Offer Quote (bps)				
	PTs			Implied RFQ	
	(1)	(2)	(3) matched to RFQ exists	(4)	(5)
ETF Deviations from NAV	0.58*** (2.93)	0.58*** (2.71)	0.54*** (2.83)	0.15 (0.59)	0.17 (0.67)
PT LCS	-	-12.63 (-0.41)	-17.16 (-0.72)	-	0.17 (1.1)
Tail Liquidity	-	61.40 (1.29)	50.78 (1.27)	-	30.7 (0.66)
Volume	-	0.03 (0.26)	-0.19 (-1.23)	-	0.02 (0.99)
Number of Line Items	-	0.10 (0.14)	-0.15** (-2.26)	-	-0.09 (-1.19)
R-squared	12.3%	15.6%	28.8%	0.8%	11.3%
PT observations	63	63	46	46	46
Sample	1st September 2022 – 26th January 2023				

Table X: Alternative Explanations – Investors "Swapping" Portfolios

The table shows the probability of an offsetting portfolio trade (i.e. in the opposite direction as the original PT) covering at least 50% of the line items in the original trade happening on the same day (T) or in the next 5 business days (T+1, T+2, T+3, T+4 and T+5).

	T	**T+1**	**T+2**	**T+3**	**T+4**	**T+5**
Mean	0.0057	0.0037	0.0039	0.0041	0.0038	0.0042
Std	0.075	0.061	0.062	0.063	0.062	0.065
P25	0	0	0	0	0	0
Median	0	0	0	0	0	0
P75	0	0	0	0	0	0
Observations	4,914 PTs executed during the period Jan 1st 2021 – Dec 31st 2021					

References

Agarwal, V., Hanouna, P., Moussawi, R., & Stahel, C. (2018). Do ETFs Increase the Commonality in Liquidity of Underlying Stocks. *SSRN Abstract 3001524*.

Amihud, Y. (2002). Illiquidity and stock returns:cross-section and time-series effects. *Journal of Financial Markets, 5*, 31-56.

Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do ETFs increase volatility? *Journal of Finance, 73*(6), 2471-2535.

Bessembinder, H. (2003). Issues in assessing trade execution costs. *Journal of Financial Markets, 6*, 233-257.

Bessembinder, H., Jacobsen, S., Maxwell, W., & Venkataraman, K. (2018). Capital commitment and illiquidity in corporate bonds. *Journal of Finance, 73*(4), 1615-1661.

Carapella, F., & Monnet, C. (2020). Dealers' insurance, market structure and liquidity. *Journal of Financial Economics, 138*(3), 725-753.

Choi, J., Kronlund, M., & Oh, J. (2022). Sitting Bucks: Zero Returns and Stale Pricing in Fixed Income Funds. *Journal of Financial Economics, 145*(2), 296-317.

Collin-Dufresne, P., Junge, B., & Trolle, B. (2020). Market Structure and Transacation Costs of Index CDSs. *The Journal of Finance, 75*(5), 2719-2763.

Corwin, W., & Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance, 67*, 719-760.

Da, Z., & Shive, S. (2018). Exchange traded funds and asset return correlations. *European Financial Management, 24*, 136-168.

Dannhauser, C., & Hoseinzade, S. (2022). The unintented consequences of corporate bond ETFs: Evidence from the taper tantrum. *The Review of Financial Studies, 35*(1), 51-90.

Dick-Nielsen, J. (2009). Liquidity Biases in TRACE. *The Journal of Fixed Income, 19*(2), 43-55.

Dick-Nielsen, J. (2014). How to Clean Enhanced TRACE Data. *Technical Report*. Retrieved from http://ssrn.com/abstract=2337908

Dick-Nielsen, J., Feldhütter, P., & Lando, D. (2012). Corporate Bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics, 103*(3), 471-492.

Edwards, A., Harris, L., & Piwowar, M. (2007). Corporate Bond Market Transaction Costs and Transparency. *The Journal of Finance, 62*(3), 1421-1451.

Ester, M., Kriegel, H.-P., & Sander, J. X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, & U. Fayyard (Ed.), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining.*

Feldhütter, P. (2012). The same bond at different prices: identifying search frictions and selling pressures. *Review of Financial Studies, 25*, 1155-1206.

Friewald, N., Jankowitsch, R., & Subrahmanyam, R. (2012). Illiquidity or credit deterioration: a study of liquidity in the US corporate bond market during financial crises. *Journal of Financial Economics, 105*(1), 18-36.

Goldberg, J., & Nozawa, Y. (2020). Liquidity Supply in the Corporate Bond Market. *Journal of Finance*.

Goldstein, I., Jiang, H., & Ng, T. D. (2017). Investor flows and fragility in corporate bond funds. *Journal of Financial Economics, 126*, 592-613.

Goldstein, M., & Hotchkiss, E. (2020). Providing liquidity in an illiquid market: Dealer behavior in US corporate bonds. *Journal of Financial Economics, 135*(1), 16-40.

Hagströmer, B. (2021). Bias in the effective bid-ask spread. *Journal of Financial Economics, 142*(1), 314-337.

Holden, C., & Nam, J. (2019). Market accessibility, Corporate bond ETFs, and Liquidity. *Available at SSRN 3083257*.

Jiang, H., Li, D., & Wang, A. (2021). Dynamic Liquidity Management by Corporate Bond Mutual Funds. *Journal of Financial and Quantitative Analysis, 56*(5), 1622-1652.

Konstantinovsky, V., Yuen Ng, K., & Phelps, B. (2016). Measuring Bond-Level Liquidity. *Journal of Portfolio Management, 42*(4), 116-128.

Koont, N., Ma, Y., Pastor, L., & Zeng, Y. (2022). Steering a Ship in Illiquid Waters: Active Management of Passive Funds. *SSRN Working Paper Abstract 4053844*. Retrieved from Available at SSRN: https://ssrn.com/abstract=4053844

Lesmond, D., Ogden, J., & Trzcinka, C. (1999). A new estimate of transaction costs. *Review of Financial Studies, 12*, 1113-1141.

Marta, T. (2020). Fixed Income ETFs, Bond Liquidity, and Stressed Markets. *Available at SSRN 3350519*.

Meli, J., & Todorova, Z. (2022). Credit ETFs in Mutual Funds and Corporate Bond Liquidity. *Working Paper*.

Pu, X. (2009). Liquidity commonality across the bond and CDS Markets. *Journal of Fixed Income, 19*, 26-39.

Rhodes, M., & Mason, J. (2022). ETF ownership and firm-specific information in corporate bond returns. *Journal of Financial Markets*.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance, 39*, 1127-1139.

Schestag, R., Schuster, P., & Uhrig-Homburg, M. (2016). Measuring liquidity in bond markets. *Review of Financial Studies, 29*, 1170-1219.

Shim, J., & Todorov, K. (2021). ETFs, illiquid assets and fire sales. *SSRN Working Paper Abstract 3886881*. Retrieved from https://ssrn.com/abstract=3886881

Ye, S. (2019). How do ETFs Affect the Liquidity of the Underlying Corporate Bonds? *Chinese University of Hong Kong Working Paper*.

Data Appendix
A1. The Machine Learning Algorithm
Step 1: Matching Barclays Inquiries to TRACE

The TRACE rules require dealers to report a trade for each individual bond in the portfolio with an attributed dollar price and an execution timestamp, despite the fact that technically the dealer and the client agree on a single price for the entire basket of bonds. This means that the individual line items must appear in TRACE if a portfolio inquiry is executed. For each line item in our portfolio inquiries database, we search through the TRACE database for Dealer-to-Customer trades which exactly match the line items in the inquiry on CUSIP, date, quantity traded and direction (dealer buy or dealer sell) [20]. The result of this process is a database of *traded* inquiries, augmented with the *exact* execution time stamp and the executed price, both of which are recorded in TRACE.

In 85% of the matches we find in TRACE, there is exactly one trade which satisfies the criteria above. The difficulty comes from the remaining 15% for which there are multiple matches. This occurs because our inquiries database only records the date but not the exact execution time stamp. Due to the enormous number of trades in TRACE, in some cases it is not possible to identify the trade without the time stamp. This applies particularly for trade sizes less than \$250K and trades executed around busy times sometimes cannot be identified without the exact execution timestamp. To determine the most likely candidate for a given bond belonging to a portfolio inquiry where multiple candidates are available, we use the distribution of the execution timestamps of the other line items in that portfolio trade. For example, we know that on Jan 5th 2021, a \$250K buy trade in bond X was part of a portfolio inquiry. Assume we find three such trades executed at 10:00, 12:30 and 14:30. If the majority

[20] We define a dealer as a traditional broker-dealer or as an alternative trading system (ATS), which we identify setting the field "Reporting Party Type" as either "D" or "T". A dealer must report a trade if the counterparty is either a customer or an affiliate, which we identify by setting the field "Contraparty Type" as either "C" or "A".

of the other line items in that inquiry were executed around 14:30, we would select the $250K trade in bond X at 14:30 and discard the other two candidate trades.

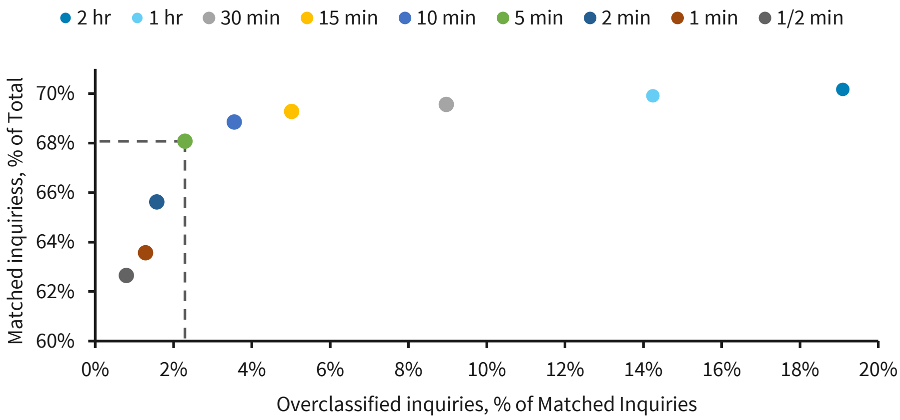
To create the blueprint of the typical portfolio trade, we need to convert the matched line items to matched portfolio inquiries. However, in doing so, we face the following trade-off. On the one hand, we want to find as many of the inquiries that actually traded as possible, but on the other, we want to minimize the number of individual bond trades we incorrectly classify as part of a portfolio i.e. the false positives. To strike the optimal balance between these goals we can pull two levers – (1) the maximum time period within which those line items must be executed; and (2) the minimum percentage of line items in the inquiry required to classify a portfolio as found[21].

To illustrate, assume that we require to find at least 80% of the line items in an inquiry. As we increase the time span between the trades that we consider, we will match more of the line items, and thus match more of the portfolios. However, we also risk over-classifying trades in TRACE, which just happen to have the same notional and the same direction as the portfolio inquiry but were not part of it. ***Figure A1.1*** demonstrates this trade-off. If we allow a time interval of 2 hours, we find 70% of the portfolio inquiries, but we over-classify 20% of the line items (i.e. 20% of the line items have multiple matches). By tightening the time interval to 5 minutes, we steeply reduce over-classification to c.2% at the cost of finding only slightly fewer of the inquiries. We conclude that 5 minutes is the optimal time interval since tightening beyond that only marginally improves precision, but drastically reduces the proportion of the inquiries that we can find.

[21] For example, if we were only able to match two line items out of one hundred included in a portfolio inquiry, and they occurred hours apart, then we clearly have not actually found the portfolio inquiry in TRACE. In contrast, if we find all one hundred line items within two seconds, then we are quite confident that we've located the portfolio trade.

Figure A1.1: Varying the Maximum Time Interval When Matching Portfolio Inquiries

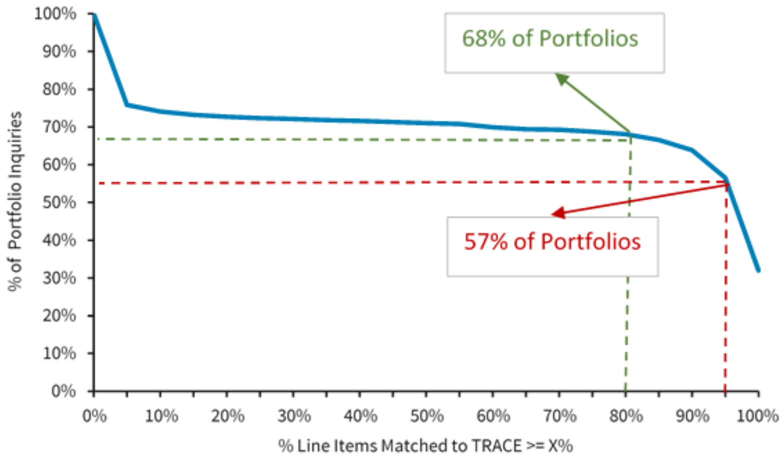
The figure shows the trade-off between the percentage of portfolio inquiries we find in TRACE against the over-classification error, as we vary the maximum time span we allow between the first and the last line item in any given portfolio inquiry.



In ***Figure A1.2*** we vary the threshold of line items per portfolio we require to match in TRACE. We find 68% of all inquiries with at least 80% of line items. In comparison, we find 57% of all inquiries with at least 95% of line items. Interestingly, the percentage of inquiries we find decreases from 100% to 75% as we just increase the threshold from 0% to 5%, but then decreases only very slowly as we further tighten the criterion. This suggests that we either find the inquiries (almost) in full or we don’t find them at all. This confirms the anecdotal evidence we have received from our conversations with portfolio desk traders about the take-it-or-leave-it nature of portfolio trades. Nonetheless, we set a rather conservative threshold of 80% of line items found in order to minimize classification error.

Figure A1.2: Varying the Minimum Number of Matched Line Items

The figure shows the percentage of portfolio inquiries we find as we vary the minimum number of matched line items per portfolio inquiry.



Step 2: Clustering algorithm

The two most important parameters of the ML algorithm we train are the maximum time we allow to elapse between the lines items in any given portfolio trade and the characteristics of the typical portfolio trade. We select and tune both parameters based on the proprietary dataset of portfolio inquiries matched to TRACE in Step 1.

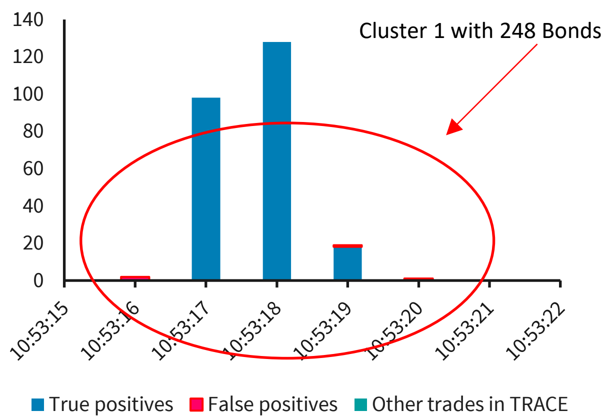
Analysing the inquiries, we discovered that when we plot the number of trades recorded in TRACE per each second of the trading day, seconds during which portfolio inquiries were executed appear like spikes or clusters (***Figure A1.3***). However, the problem is that such clusters are both rare (compared to the total volume that appears on TRACE) and could take very different time to appear on the TRACE tape. For example, most inquiries span zero to two seconds, but some of the larger ones could take up to 20 seconds. This means that we need to develop an algorithm which is able to separate the large amount of "noise" in the data (i.e. the non-portfolio trades), but is flexible enough to accommodate different portfolio structures. In other words, the algorithm needs to allow for different length (in terms of time) of the portfolios. For example, an algorithm which identifies clusters based on a fixed time interval, no matter how tight that interval is, would produce noisier estimates.

We employ a machine learning toolkit and use a DBSCAN clustering algorithm (Density-based Spatial Clustering of Applications with Noise) to obtain a list of portfolio *candidates* (Ester, Kriegel, & Sander, 1996). DBSCAN searches through the millions of TRACE observations and forms clusters of trades whose execution timestamps are closely packed together (i.e. the trades have many nearby neighbours) and marks as outliers points that are located in low-density regions (i.e. their nearest neighbours are too far away). Clusters identified in this way are strictly non-overlapping and the individual line items included in any cluster are unique to that cluster only.

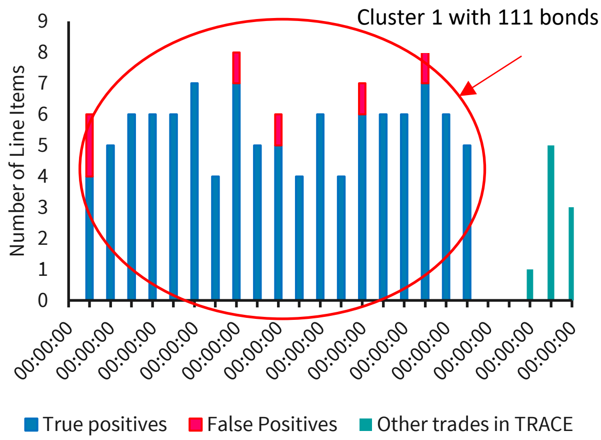
More specifically, each day from January 1st 2018 to December 31st 2021, the algorithm orders all dealer-to-customer trades in TRACE by their timestamp and, starting from the first trade on that day, searches for trades which have at least 25 other trades recorded within a two second interval. Each such a trade is labelled as a “core” trade. Some of the trades that are within the two second neighbourhood of a “core” trade could be “core” trades themselves. For example, if a trade that is exactly two seconds from the original “core” trade has at least 25 trades within its own two second window, it too would be a “core” trade. We then link “core” trades and their two second neighbourhoods to form a cluster. In other words, each cluster must contain at least one core point. Further, individual trades in a cluster may well be more than two seconds apart, but *any* trade in a cluster is *at most* two seconds away from some core trade. It is precisely the “expanding” nature of the algorithm, which produces clusters with different time length.

Figure A1.3: Examples of How Portfolio Trades Appear in TRACE

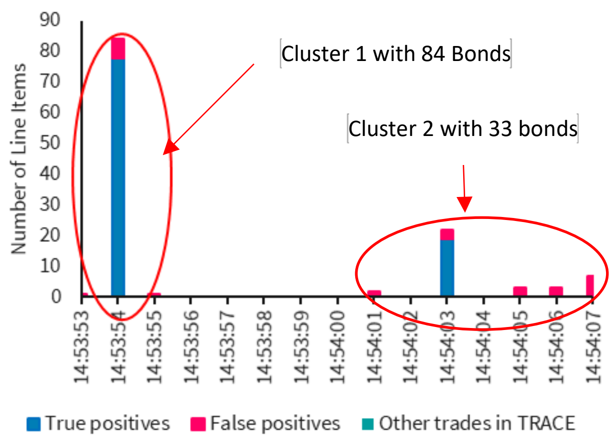
Example 1 "Tight" – Inquiry with 244 Bonds



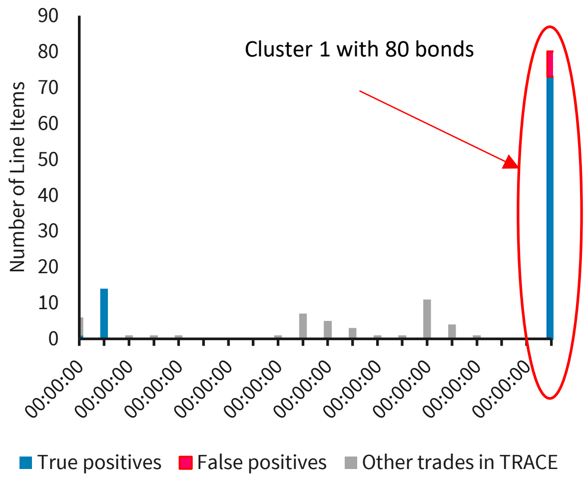
Example 2 "Spread-out" – Inquiry with 105 Bonds



Example 3 "Batched" – Inquiry with 98 Bonds



Example 4 "Batched" – Inquiry with 88 Bonds



Step 3: Re-clustering

Each of the clusters the algorithm identifies has a high probability of being an actual portfolio trade. As shown by the examples above, some portfolio trades are split into multiple batches, and others are in one, and it is extremely difficult for our algorithm to tell which is which. This is why we think this approach will give accurate estimates of the volumes associated with this trend, but less accurate estimates of the boundaries of portfolio trades, and hence the overall count. However, since we eventually want to test how execution quality differs across different portfolio construction strategies, it would be extremely valuable to reconstitute these clusters into their original portfolios, if possible.

To this end, we re-classify the clusters from the previous step by aggregating those clusters that happen one minute apart into a unified portfolio. The idea is to bring together several batches of the same portfolio (as in Example 3 and 4 on ***Figure A1.3***). It is important to note that we neither add nor delete portfolio volume in this step – we simply adjust the boundaries of the clusters.

Step 4: Filtering

Next, we filter this list to remove candidate clusters than don't line up with what we expect given the analysis of our inquiry in Step 1:

- We drop clusters that are within 5-minute intervals before and after popular delayed spot times – 11.00, 15.00, 15.30, 16.00, 16.30. As a result, we are likely to understate the true prevalence of portfolio trades because some IG portfolio trades are certainly spotted at these times. However, if we don't drop those clusters we are certain to include lots of false positives.
- We drop clusters that contain less than \$5 million in HY and \$10 million in IG, and clusters with average line item size below \$100K in HY and below \$250K in IG. This is necessary to reduce false positives associated with odd lots trading, much of which is electronic.

Finally, we fine-tune by deleting a small number of line items which are markedly different from the cluster to which they belong. These adjustments have a minor impact on the total portfolio volume we identify but substantially reduce the portfolio-level false positives rate:

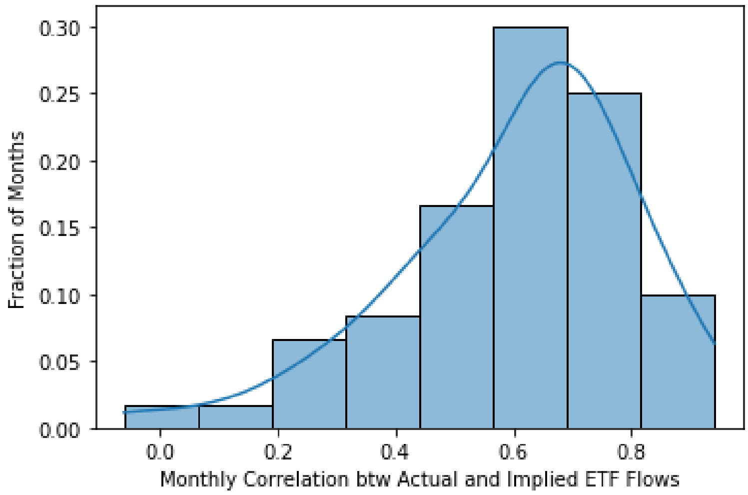
- We know from our inquiries that portfolio trades are either buy-only, sell-only or balanced buy-and-sells trades. For example, if we see a candidate cluster with 100 bonds, the most likely distribution of trades is – a client buys 100 bonds; a client sells 100 bonds and client buys 50 bonds and sells 50 bonds. Hence, a candidate cluster where a client buys 97 bonds and sells 3 bonds is extremely unlikely – in reality this is a buy-only trade with 97 bonds and the 3 sell trades were coincidentally executed at the same time.

- Similarly, the majority of portfolio inquiries are HY only or IG only. Whenever HY and IG bonds are mixed in the same portfolio trade, these are likely to be at the boundary between HY and IG – e.g. a mix of BAA3s and BA1s. In other words, a candidate cluster of 95 B2 bonds and 5 AAA bonds is highly unlikely, even if the direction (buy or sell) matches. This is likely to be a straight HY trade with 95 bonds.

A2. Figures

Figure A2. 1: Correlation Between Actual and Implied ETF Flows

The figure shows the histogram of monthly correlation coefficients between actual and implied LQD flows. We obtain actual flows data from Bloomberg. Implied flows data are estimated following the procedure developed by Shim and Todorov (2021) and Koont et al. (2022) (for details refer to Section 2.3 ETF Sample). Estimation period January 2018-December 2022.



A3. Tables

Table A3.1: Other Portfolio Strategies

The table shows a summary of the different strategies investors use when trading portfolios of bonds. When classifying a portfolio trade we always compare it to the Bloomberg IG Corporate Bond Index. A *Liquidity* strategy applies in those cases where the portfolio trade contains at least 50% more trade volume in the 4th or 5th LCS quintile (most illiquid quintiles) than what we would normally expect in the IG Index. A *Market View* strategy applies in those cases when the portfolio maturity/sector/rating Herfindahl score (HHI) is at least 50% higher than the respective HHI of the Index. Portfolio trades that are neither axed towards a *Liquidity* nor a *Market View* strategy are classified as *Diversified*.

Type of Strategy	**% of IG PT Volume**
➢ Liquidity	49%
➢ Market View	
o Maturity View	35%
o Sector View	24%
o Rating View	13%
Concentrated (Liquidity OR Market View)	**69%**
Diversified (Flows Management)	**31%**

Table A3.2: Robustness – Fixed Effects

$\boldsymbol{EHS_{i,j,t} = \beta_1 Port\ Trade_{i,j,t} + \Gamma Z_{j,t} + \delta Block\ Trade_{i,j,t} + \Theta Y_y + \lambda_{j,t} + \epsilon_{i,j,t}}$

The table reports transaction-level regressions of effective half spread on a portfolio trade dummy ($Port\ Trade_{i,j,t}$) and a set of controls. Bond-level controls include: maturity, time since issuance, numeric rating, option-adjusted spread, logarithm of issue size, coupon, call type and sector dummies and time until next call. Date-level controls include: the VIX, the logarithm of total trading volume and percentage of block trades (>$5 million). T-stats in parentheses. Standard errors are clustered at the bond and date (day) level. Significance at the 1 %, 5 % and 10 % statistical level is denoted by ***, **, and * respectively.

	Effective Half Spread (EHS)			
	TRACE PT		Client Inquiries	
	(1)	(2)	(3)	(4)
Portfolio Trade	-6.84*** (-24.66)	-8.72*** (-33.93)	-5.99*** (-4.76)	-10.44*** (-5.67)
Bond-level Controls	YES	YES	YES	YES
Trade-level Controls	YES	YES	YES	YES
Date-level Controls	NO	YES	NO	YES
Bond FE	NO	NO	NO	NO
Date FE	NO	NO	NO	NO
Bond-Date FE	YES	NO	YES	NO
Bond-trade Observations	4,467,324	3,818,560	4,467,324	3,818,560
Sample	Bonds included in LQD Jan 1st 2021 – December 31st 2021			

Table A3.3: Relationship to the ETF C/R – "Right way" for the ETF

$$\boldsymbol{EHS_{i,j,t} = \beta_1 Port\,Trade_{i,j,t} + \beta_2 Right\,way^{LQD}_{i,j,t} + \beta_3 Port\,Trade_{i,j,t} \times Right\,way^{LQD}_{i,j,t} + \Gamma Z_{j,t} + \delta Block\,Trade_{i,j,t} + \lambda_j + \delta_t + \epsilon_{i,j,t}}$$

The table reports transaction-level regressions of effective half spread on a portfolio trade dummy ($Port\ Trade_{i,j,t}$), a trade-level dummy variable ($Right\ way_{i,j,t}$) and their interactions for all bonds owned by the largest IG ETF (ticker LQD). $Right\ way_{i,j,t}$ equal 1 for all customer sell (buy) trades on days when LQD is mostly creating (redeeming) shares and 0 otherwise. We split our sample into three sub-samples using terciles of the bond-level LCS distribution: liquid, medium and illiquid. Regressions control for bond-level maturity and numeric rating (higher values are worse) (collected in vector $Z_{j,t}$) and include a transaction-level dummy variable equal to 1 for all trades larger than \$5 million ($Block\ Trade_{i,j,t}$). All variables are winsorized at the 1% level. All regressions include bond and date fixed effects. T-stats in parentheses. Standard errors are clustered at the bond and date (day) level. Significance at the 1 %, 5 % and 10 % statistical level is denoted by ***, **, and * respectively.

	Effective Half Spread (EHS)			
	(1) All LQD Bonds	(2) Liquid	(3) Medium	(4) Illiquid
Portfolio Trade	-6.33*** (-10.29)	-5.16*** (-17.37)	-6.75*** (-11.11)	-7.78*** (-5.17)
Right way	0.86 (0.51)	1.11 (1.14)	0.58 (0.31)	0.51 (0.13)
Portfolio Trade × Right way	-1.67* (-1.76)	-0.91** (-1.95)	-1.27* (-1.69)	-3.39* (-1.75)
Maturity	0.55*** (11.82)	-0.10 (-0.63)	0.51*** (4.20)	2.29*** (4.44)
Numeric rating	0.35 (1.38)	0.14 (0.51)	-0.44 (-0.85)	2.88*** (2.38)
Block Trade	-3.43*** (-10.40)	3.28*** (-16.38)	-4.34*** (-11.36)	-2.47*** (-3.39)
Bond FE	YES	YES	YES	YES
Date FE	YES	YES	YES	YES
Bond-trade Observations	2,324,395	1,088,002	837,440	398,953
Sample	Bonds included in LQD Jan 1st 2021 – December 31st 2021			

Table A3.4: Relationship to the ETF C/R – "Right way" for the ETF with Portfolio Inquiries

$$\boldsymbol{EHS_{i,j,t} = \beta_1 Port\,Trade_{i,j,t} + \beta_2 Right\,way^{LQD}_{i,j,t} + \beta_3 Port\,Trade_{i,j,t} \times Right\,way^{LQD}_{i,j,t} + \Gamma Z_{j,t} + \delta Block\,Trade_{i,j,t} + \lambda_j + \delta_t + \epsilon_{i,j,t}}$$

The table reports transaction-level regressions of effective half spread on a portfolio trade dummy ($Port\ Trade_{i,j,t}$), a trade-level dummy variable ($Right\ way_{i,j,t}$) and their interactions for all bonds owned by the largest IG ETF (ticker LQD). $Right\ way_{i,j,t}$ equal 1 for all customer sell (buy) trades on days when LQD is mostly creating (redeeming) shares and 0 otherwise. We split our sample into three sub-samples using terciles of the bond-level LCS distribution – liquid, medium and illiquid. Regressions control for bond-level maturity and numeric rating (higher values are worse) (collected in vector $Z_{j,t}$) and include a transaction-level dummy variable equal to 1 for all trades larger than \$5 million ($Block\ Trade_{i,j,t}$). All variables are winsorized at the 1% level. All regressions include bond and date fixed effects. T-stats in parentheses. Standard errors are clustered at the bond and date (day) level. Significance at the 1 %, 5 % and 10 % statistical level is denoted by ***, **, and * respectively.

	All LQD Bonds	Liquid	Medium	Illiquid
Portfolio Trade	-5.16** (-1.96)	-4.60*** (-3.02)	-5.08** (-1.97)	-5.52* (-1.79)
Right way	0.72 (0.43)	0.11 (1.09)	0.47 (0.26)	0.12 (0.03)
Portfolio Trade × Right way	-2.21 (-0.47)	-0.94 (-0.45)	-2.02 (-0.48)	-3.37 (-0.44)
Maturity	0.51*** (11.07)	-0.13 (-0.77)	0.48*** (3.97)	2.25*** (4.43)
Numeric rating	0.36** (1.69)	0.13 (0.51)	-0.45 (-0.87)	3.04** (2.43)
Block Trade	-2.94*** (-8.97)	-3.01*** (-15.02)	-3.82*** (-9.98)	-1.56** (-2.17)
Bond FE	YES	YES	YES	YES
Date FE	YES	YES	YES	YES
Bond-trade Observations	2,324,395	1,088,002	837,440	398,953
Sample	Bonds included in LQD Jan 1st 2021 – December 31st 2021			