Online Appendix for

The Effects of Mandatory Transparency in Financial Market Design: Evidence from the Corporate Bond Market[1]

Paul Asquith Thomas R. Covert Parag A. Pathak[2]

This draft: April 4, 2019

[1] We are grateful to Hank Bessembinder, Edith Hotchkiss, Leonid Kogan, Deborah Lucas, Jun Pan, and Alp Simsek for discussions, and Ola Persson and FINRA for conversations about the data. We also thank Jane Choi, Abhishek Dev, Daniel Green, Joseph Shayani, Josh Weiss, and Ahmad Zia Wahdat for their research assistance.

[2] Asquith: MIT Sloan School of Management, Cambridge, MA 02142 and NBER, email: pasquith@mit.edu; Covert: Chicago Booth School of Business, Chicago, IL 60637, email: Thomas.Covert@ChicagoBooth.edu; Pathak: MIT Economics, Cambridge, MA 02142 and NBER, email: ppathak@mit.edu.

Appendix B: TRACE Data

Our data source for corporate bond trading is the Academic Corporate Bond TRACE data, i.e. Academic TRACE, purchased from FINRA. Beginning in July 2002, TRACE publicly provided price and volume for large, highly rated bonds in disseminated trades. FINRA classified these first bonds as Phase 1 bonds and over the course of the next three years, FINRA released trading data for all corporate bonds in a total of four phases. Throughout, exact volumes were reported for trades less than 1MM for high-yield bonds and trades less than 5MM for investment grade bonds. Amounts over that were only reported as greater than 1MM and/or 5MM, respectively.

In March 2010, FINRA released a "Historical" TRACE dataset, which includes both disseminated and non-disseminated trade reports, starting from TRACE's initiation in July 2002, and also includes the exact volume for all trades. That is, it released trade reports for bonds that had not yet been "phased in" and which were not publicly disseminated at the time of the trade. In February 2017, FINRA appended anonymized dealer identifiers to this Historical TRACE dataset to form the Academic Corporate Bond TRACE dataset.

During our time period, July 1, 2002 until December 31, 2006, there are 37,934,432 unique trade reports on 41,676 different CUSIPs in the Academic TRACE dataset. FINRA requires that all FINRA registered dealers self-report trade reports. This means that an interdealer trade should be reported twice, by both the buying dealer and the selling dealer. A trade between a dealer and customer is only reported once, by the dealer.

There are a number of reporting errors in this self-reported data. This appendix and the corresponding Tables B1 and B2 describe the steps we take to convert the raw Academic TRACE dataset to our analysis sample. That is, we list the steps we took to eliminate duplicate trade reports for the same trade and to correct errors in reporting.

We use the following fields, defined below, in the data set to construct our cleaned sample:

- **First Trade Control Number** is the unique (within the reporting day) number assigned by TRACE to a trade report
- **CUSIP** is a universal security identifier
- **Execution Date** is the date the trade was reported executed
- **Execution Time** is the time the trade was reported executed
- **Reporting Date** is the date the trade was actually reported to TRACE
- **Reporting Time** is the time the trade was actually reported to TRACE
- **Price** is the reported bond price and excludes any commission reported by the firm in the transaction[3]

[3] We received one of the earliest releases of the Academic Corporate Bond database. The version of the data that FINRA originally provided to us includes a field, CMSN_TRD, which purports to indicate whether a commission is assessed on the trade. The subsequent version of the dataset that we use does not include this information. In the TRACE User guide (2008 and 2017 versions), it states that "The reported price for a reported transaction must

- **Quantity** is the reported volume as a number of bonds
- **Reporter ID** is the anonymized hash code for the dealer who reports the trade
- **Give-up Reporter ID** is the hash code for the dealer in Give-up trades
- **Contraparty ID** is the hash code reported for the party on the other side of the trade
- **Give-up Contraparty ID** is the hash code for the party on the other side of the trade in Give-up trades
- **Reporter Capacity** indicates whether the reporter is acting as an agent or principal
- **Contraparty Capacity** indicates whether the contraparty is acting as an agent or principal
- **Buy/Sell Indicator** reports whether the reporting party is buying or selling
- **Dissemination** reports whether a trade was disseminated or not
- **Locked-in Flag** indicates that the reporting party submits the trade report on its own behalf as well as on behalf of the contraparty
- **Trade Status** corresponds to the type of report (trade, cancellation, modify)
- **As Of Indicator** represents if a trade is reported for a prior business day as in the case of a reversal or when the execution and reporting date differ

Most of these fields are self-reported. More detailed definitions for some of the variables will appear below as we discuss their use. There are additional fields associated with each trade report, which we do not use as part of our analysis.

Eliminate bonds based on characteristics

Before eliminating and correcting trade reports, we match the Academic TRACE dataset to the universe of corporate bonds in the Mergent FISD database. FISD is our source for bond characteristics such as issue size, ratings, maturity, etc. which we add to the Academic TRACE dataset. We drop all TRACE bonds that do not match to FISD by CUSIP. We also drop all bonds with equity-like characteristics (convertibles, exchangeables, etc.) since their equity component may be included in the bond price. We next drop all Rule 144a bonds because TRACE did not disseminate trading information on these bonds during 2002-2006. Finally, FISD does not report a correct issue size in some cases. For example, there are some bonds in FISD with reported issue size of $0. After hand checking a number of cases with small issue size, we decided to drop all bonds with reported FISD issue size of less than $100,000. The number of trades eliminated and their corresponding CUSIPs affected by these steps are shown in the first section of Table B1. In total, we drop 10,018 CUSIPs which correspond to 4,652,061 trade reports.

Eliminate trade reports because of self-reported errors

be inclusive of a mark-up or markdown, but is exclusive of a commission, since a commission is reported in a separate field. TRACE will, however, incorporate the commission into the disseminated price." In the TRACE data code book (before 2012 and the Academic Corporate Bond database from 2017), it says: "This field [price] represents the reported bond price and is inclusive of any commission, mark-ups, and/or mark-downs reported by the firm in the trade transaction." Based on our analysis, we believe that reported price in our dataset does not include commissions.

Next, we eliminate trade reports which do not take place as reported since they are later modified, cancelled, or reversed. These are catalogued in the second section of Table B1. These are trade reports where FINRA dealers have themselves identified errors. That is, FINRA dealers insert a second trade report to correct a previous report. When identifying trades reports that are subsequently modified, cancelled or reversed, we rely on two data fields: Trade Status and First Trade Control Number.

The field Trade Status can take on values of "T" (new trade report), "X" (trade cancel), "C" (correction cancel), "R" (new correction), and "Y" (reversal). When a trade is first reported, it has a trade status of "T." If a member firm wishes to cancel a trade they reported earlier in the day, it must submit a new report containing the same CUSIP, Date, Price, Quantity, Reporter ID, and Contraparty ID of the original trade and mark Trade Status with an "X", indicating that this trade is cancelling the original trade. If a member firm wishes to correct a trade reported earlier in the day, it must submit two reports. One report, containing the details in the original report, must be marked "C" to indicate that the original report is cancelled as part of a correction. The other report, containing the correct trade information, must be marked "R" to indicate that it is a newly reported, correct trade. Cancellations that are reported on a different day than the original report are categorized as reversals and marked "Y".

The field, First Trade Control Number, facilitates tracking corrections and cancellations. This field contains a unique number for each new trade report. Any report that is cancelling or correcting an earlier report will contain the First Trade Control Number of the original report so that it can be linked back to the proper report. We group together reports with the same First Trade Control Number and order reports by reporting time within each group to determine which trades have been cancelled or corrected and thus need to be eliminated. The last report within a group indicates which reports to eliminate. For example, if the last report is "R" (new correction), then the last report will be kept, and all earlier reports will be eliminated. If the last report is "X", then all reports in the group will be eliminated as this trade has been cancelled. In our dataset, all modification and cancellation reports can be matched to an original trade report as Table B1 shows. We eliminate 1,250,351 reports due to modifications and 1,013,223 reports due to cancelations, including both the original trade reports and the modify/cancel reports.

The last Trade Status is reversals, which are cancellations that are reported on a different day from the original trade report. Trade reversals are identified by a "Y" in the Trade Status field and also by the As Of Indicator field by "R". Since the original trade and its reversal are reported to TRACE on different days, the First Trade Control Number of reversal trades do not necessarily match the First Trade Control Number of the original trade. Therefore, to link a reversal to its original trade, we match the reports using nine identifying characteristics: CUSIP, Execution Date, Execution Time, Price, Quantity, Reporter ID, Contraparty ID, Buy/Sell indicator, and Reporter Capacity. We called matches with these criteria a "nine-way" match.

Using these nine characteristics sometimes leads to many-to-many matches; that is, there is often more than one possible pairing. (In fact, it appears that many reversals are the result of a trade being entered twice and the second report being reversed). After matching reversal and non-reversal

reports using these nine trade characteristics, we eliminate exact matches as follows. If there is only one exact match, both the reversal and its matched trade are eliminated. If there is more than one exact match, we eliminate the reversal trade and one of the matching trades. Since, by definition, the trades occur at the same time, date, price, and volume, the cleaned dataset is unaffected by the choice of which matching trade reports we eliminate. For example, if there are four reversals and five non-reversals, we drop the four reversals and drop the first four non-reversals. We drop a total of 841,108 trade reports due to exact matches of reversal reports, most of which are not part of many-to-many matches.

Unfortunately, unlike all modifications and cancellations, not all reversals have an exact nine-way match. A large number of the unmatched reversals have an eight-way match to another trade if we drop the same execution time requirement. Since execution time is self-reported, we assume these eight-way matches were the original trades that were meant to be reversed, and we eliminate the reversal and the matched trade from the sample following the steps above. We drop an additional 87,250 reports. Next, we relax the requirement that the price must be exact in an eight-way reversal and look for matches when prices are rounded to 0.01.[4] This relaxation allows us to eliminate 5,694 additional trade reports. Even after eight-way matches, there are 30,724 reports labeled as reversals that we were unable to match to an original trade. We dropped these reversal reports from the dataset.

Table B1 shows that 3,228,350 trade reports are eliminated from the sample because of self-reported errors.

Modifying trade reports based on TRACE submission rules

There are multiple conventions or paths by which the transacting parties can be reported to TRACE. In addition to self-reporting, TRACE also allows another party (such as a clearing firm) to fulfill the reporting obligation of transacting party. There are three fields that are used when trade reports do not report the transacting dealers in the trade. They are Reporter Give-up ID, Contraparty Give-up ID, and Locked-in Flag. In a Give-up trade, a clearing firm can submit a report on behalf of the either of the transacting dealers. When this is the case, either the Reporter Give-up ID or Contraparty Give-up ID is populated by the ID of the transacting dealer. To correctly identify the transacting parties in these trades, we replace the Reporter ID (Contraparty ID) with the Reporter Give-up ID (Contraparty Give-up ID).

The other case where TRACE allows a variance on its reporting requirements is a locked-in trade report. In a locked-in trade report, the reporting party submits the trade report on its own behalf as well as on behalf of the contraparty. That is, there is only one trade report, rather than a separate report from the buying dealer and selling dealer. When the Locked-in Flag is checked, the reporter ID and the contraparty ID are the same, but one give-up field is populated. For each trade report where the Locked-in Flag is marked, we follow the convention in the paragraph above for the give-up fields and modify the Reporter or Contraparty ID fields in the existing trade report. In addition, we create a new

[4] Prices are only rounded in this step to identify reversals. In subsequent steps, we return to the original price.

buy or sell trade report corresponding to the missing report. For this missing report, we treat the contraparty of the locked-in report as the reporting party and use the Contraparty Capacity field of the locked-in report for the Reporting Capacity field. It is important to create these extra reports for the matching process described below. Since these newly created reports correspond to the same trade as an existing report, they will be eliminated following the steps to identify duplicate trade reports. Table B1 shows that 523,125 extra trade reports are created.

Eliminate trade reports with timing issues

We next drop trade reports from our sample because of issues with their execution date. We eliminate a small number of trade reports reported to be executed before July 1, 2002, which is the beginning of TRACE. Then, we drop any trade report that occurs before the traded bond's offering date or after its maturity date. Many trades that occur before the offering date correspond to the when-issued market. Trades that occur after maturity date often occur on bonds that are bankrupt. Finally, we drop all reports that are reported to have occurred on SIFMA holidays. Even though bond trades occur on these holidays, we drop these days because trading activity is limited to a subset of dealers. After these eliminations, we are left with 29,387,759 reports.

Eliminate redundant trade reports

A bond trade may generate multiple trade reports. To identify redundant trade reports, we define a bond trade as an exchange of bonds between two principals. We define principals as parties (dealers or customers) that hold a bond for some time period and in so doing assume the risk that the bond price may change. A significant part of our data cleaning process involves identifying and keeping bond trades while eliminating extraneous trade reports.

FINRA requires that every dealer in a trade report a separate trade report. (As mentioned above, customer do not report trade reports.) This means for example, if two dealers trade a bond between them, both dealers will file a trade report. Dealer A will report they sold a bond to Dealer B, while Dealer B will report they purchased a bond from Dealer A. Since these two reports only represent one trade, it is important to identify that fact and consolidate the two trade reports into one.

In some trades, another dealer (or more) may act as an agent. In an agency transaction, a dealer intermediates between two principals by transferring the bond from one principal to the other while not assuming any price risk. We will refer to the trade reports that involve agency transactions as an agency chain. In a common case, Dealer C will purchase a bond from Dealer A and transfer it to Dealer B at the same price plus a commission. (Commissions are reported in a separate field and are not used to eliminate redundant reports.) In this case, TRACE requires four trade reports to be generated: Dealer C reports she bought the bond from Dealer A as an agent and sold the bond to Dealer B as an agent. Dealer A reports she sold the bond to Dealer C as a principal and Dealer B reports she bought the bond from Dealer C as a principal. Thus, this one bond trade will generate four trade reports, which must be identified and culled to one report in order to properly measure trading activity.

For all trades, our aim is to keep only one trade report that reflects the transfer of bond ownership and price risk. We next present some more formal examples of the discussion above. An agency chain involving three dealers generates four trade reports in the following way:

Reporter	Contraparty	Buy/Sell Indicator	Reporter Capacity
Dealer A	Dealer C	Sell	Principal
Dealer C	Dealer A	Buy	Agent
Dealer C	Dealer B	Sell	Agent
Dealer B	Dealer C	Buy	Principal

In this example, Dealer C acts in an agency capacity to intermediate between two principals: Dealer A and B. Therefore, we consolidate these four trade reports into a single trade record as follows.

Reporter	Contraparty	Buy/Sell Indicator
Dealer A	Dealer B	Sell

This consolidated record never shows up as a report; we construct it from the four trade reports. We mark this consolidated trade record as an agency chain with Dealer C as agent and adopt the convention of preserving the sell-side report.

The most common type of agency chain occurs when a customer is involved as a principal (either as the final Buyer or initial Seller) as illustrated below:

Reporter	Contraparty	Buy/Sell Indicator	Reporter Capacity
Dealer A	Customer	Buy	Agency
Dealer A	Dealer B	Sell	Agency
Dealer B	Dealer A	Buy	Principal

In this example, Dealer A acts in an agency capacity to intermediate between a customer and Dealer B. We consolidate these three trade reports into a single agency trade record as follows.

Reporter	Contraparty	Buy/Sell Indicator
Dealer B	Customer	Buy

Agency chains can be far more complicated than these examples and involve longer sequences of trade reports where the reporter capacity is Agency. We discuss this below.

Turning now to principal-to-principal trades, the simplest example of a principal-to-principal trade with redundant reports is the following:

Reporter	Contraparty	Buy/Sell Indicator	Reporter Capacity
Dealer A	Dealer B	Buy	Principal

Dealer B	Dealer A	Sell	Principal

We consolidate these trade reports into a single record as follows:

Reporter	**Contraparty**	**Buy/Sell Indicator**
Dealer B	Dealer A	Sell

To identify candidate redundant reports, for both agency and principal trade reports, we match trade reports on the following fields: CUSIP, Execution Date, Execution Time, Price, Quantity, Reporter ID, Contraparty ID, Buy/Sell Indicator, and Reporter Capacity. Execution Time is used to break ties when there are many-to-many matches. Every interdealer trade should have two paired reports and every agency buy trade should be matched with an agency sell trade.

There are many instances in which a trade report cannot be matched with another one, despite that fact that a match should exist by TRACE convention. In many of these instances, potential matches differ by one field.

After working with the data, we have identified four common difficulties with matching trade reports.

1. **Trade splitting**: We find many instances in which trade reports can be matched if we sum the volume across multiple reports on either the buy or sell side of the trade. In these cases, we believe the two dealers see the trading activity differently. For example, one dealer executes a sale of volume 5,000,000 and reports it as such. Meanwhile, the other dealer reports five buy transactions of 1,000,000.
2. **Reporting capacity mismatches**: There are many trades where the Reporter Capacity code appears to be inconsistent. That is, a dealer reports she is acting as an agent on one side of a transaction and a principal on the other.
3. **Dealer identification**: Many unmatched trade reports can be matched if we allow for multiple dealer IDs. This can occur, for example, if one dealer has merged with another dealer and now has two dealer IDs.
4. **Price corrections**: Many unmatched trade reports can be matched using different rounding or truncations for the per-bond price.

Addressing these difficulties by modifying the four fields: Quantity, Reporter Capacity, Contraparty ID, and Price, allows us to match many, although not all, of the unmatched reports. Modifying more fields or even multiple fields would increase our number of matches, but many of those modifications would also result in false matches. As a result, we try to minimize the number of fields we change to identify matches and only modify in a trade report one of the four fields we list above. In the following section, we describe our procedure for eliminating redundant reports in more detail.

Eliminating redundant trade reports

First, we match trade reports in agency chains without making any modifications. As shown in the examples above, one trade generates multiple trade reports which must be consolidated into one. If there is a sequence of trade reports that can be explained as an agency chain in different ways, we chose the way that corresponds to the most frequently observed type of agency chain. If there are multiple possible trade records that can rationalize a group of trade reports, we match trades chronologically by execution time. We eliminate 1,290,607 trade reports through the exact matching of agency chains.

After removing these trade reports, we next address trade splitting, wherein the dealers report one trade in multiple pieces. To deal with trade splitting, we aggregate all of the reports with the same CUSIP, Execution Date, Price, Reporter ID, and Contraparty ID. With these aggregated trade reports, we then match any additional agency chains, and we eliminate an additional 57,314 reports.

Next, we address the mismatches of reporting capacity codes. In each trade report, the reporting party classifies herself as either a principal or an agent. Every trade or agency chain must have a principal on both ends. Likewise, an intermediate dealer in an agency chain should report both an agency buy and an agency sell. We find many "broken chains" where the two trade reports from the same dealer for the same price and quantity on the same day are inconsistent. In one trade report, the dealer reports she is acting as an agent, and in the matched trade report, she reports she is acting as a principal. In these cases, we convert one of the reporting capacity codes to fix the broken agency chain. Table B1 shows that we eliminate 176,744 trade reports as a result of the new matches from modifying the capacity code. We eliminate an additional 3,736 trade reports when we aggregate volume, as above.

Then we address misreported dealer IDs. Ideally, matched buy and sell reports will have the reporter's ID of the buy report be the contraparty's ID in the sell report, and vice versa. However, a given dealer may be associated with more than one ID because of mergers, reorganizations, or because two dealers are subsidiaries of the same firm. If the reporting dealer populates the Contraparty ID field with an alternative dealer ID, we do not have an exact match. The example below illustrates this case:

Reporter	**Contraparty**	**Buy/Sell Indicator**
Dealer A	Dealer B	Buy
Dealer B	Dealer C	Sell

Suppose that dealer IDs A and C both represent the same dealer. By substituting A for C in the sell report, we can identify the match and eliminate one trade report. When we consolidate a match with a misreported contraparty ID, of the two possible dealer IDs, we always keep the ID that appears on the in the Reporter ID field. This is because we assume that the reporting side dealer's information about its ID is more likely to be accurate than the contraparty. The consolidated record is:

Reporter	**Contraparty**	**Buy/Sell Indicator**
Dealer B	Dealer A	Sell

As with Reporter Capacity mismatches, we look for misreported dealer ID matches with both exact and aggregated trade volumes. In total, modified dealer ID matches allow us to eliminate an additional 105,828 trade reports.

After eliminating redundant trade reports that are part of agency chains, we next match duplicate reports from interdealer principal-to-principal trades. Table B1 shows that 3,713,491 trade reports are dropped without any modifications to trade reports. As above, if there are multiple possible trade records that can rationalize a group of trade reports, we match trades chronologically by execution time. Then, we correct the same errors– reporting capacity mismatches and misreported dealer IDs, both with exact and aggregate volume – for these trade reports. An additional 934,536 trade reports are dropped using these modifications.

Because of the multiple dealer IDs discussed above, it is also possible that in some trades both dealers will be identified with two separate dealer IDs. For example, suppose that IDs A and C are identified with one dealer, and IDs B and D are identified with another dealer. The following two trade reports

Reporter	**Contraparty**	**Buy/Sell Indicator**
Dealer A	Dealer B	Buy
Dealer D	Dealer C	Sell

are redundant reports of a single principal-to-principal trade but will not be identified as such.

In keeping with our principle of making minimal corrections, we do not want to match these reports unless we have strong evidence that either IDs A and C or B and D belong to the same dealer. For example, if A and C represent the same dealer, then Dealer D was not really making an error when she filled in C for the contraparty.

To determine whether two IDs possibly represent the same dealer, we count how many times one ID is substituted for another in contraparty mismatches from the first pass, i.e., all the previous matching steps described above. If an ID (C, in our example) is modified at least 100 times in the first pass and the same substituting ID (A) is used in at least 95% of these instances, then we call this substitution (C -> A) a "common substitution," and we infer that the two IDs belong to the same dealer. We identify 76 common substitutions.

We make all common substitutions throughout the dataset. In our example, we replace every instance of C with A. Then we make a second pass through the remaining trade reports to identify agency chains and principal-to-principal trades following the steps we used with the original IDs. We eliminate 21,188 additional trade reports.

Next, we consider possible corrections to the price field for the remaining unmatched trade reports. First, we round the prices of these trade reports to two significant digits and make a third pass to match trade reports in agency chains and principal-to-principal trades. This results in an additional

460,405 trade reports being eliminated. We make a fourth pass with truncated prices (e.g., we truncate 100.116 to 100.11, as opposed to rounding to 100.12). We eliminate additional 26,269 trade reports.

Modify or eliminate trade reports with price or volume issues

Some trade prices are vastly out of line with other prices for the bond during the same period. To identify these prices, we first define two reference prices, before and after the trade report. The reference prices are the median prices, weighted by volume, of the 10 trades immediately before and 10 trades immediately after the trade. If there are fewer than 10 trades in the 30 days before or after, we use all the trades in the 30 days. A trade price is vastly out of line if the bond price differs by more than $20 per bond in the same direction from both the before and after reference prices. When a trade is vastly out of line, we replace it with the volume-weighted median prices for the entire 20 trades or 60-day window. A total of 6,850 prices were replaced with the reference price.

These modifications do not necessarily correct all the prices that may be erroneous. For example, the procedure above does not correct the price if there is not another trade report in the 60-day window. We, therefore, drop trade reports less than the 0.01 percentile and 99.99 percentile of all trade prices in the sample. Table B1 shows we drop 4,848 trade reports.

Next, we eliminate trades from the dataset when volume information appears erroneous. Since, as mentioned above, all trades are self-reported, data entry errors are possible even though TRACE monitors reported trades. We, therefore, drop trade reports less than the 0.01 percentile and 99.99 percentile of all trade volumes in the sample. Table B1 shows we drop 3,283 trade report. We also eliminate 4,396 trade reports in which the volume of a single trade is higher than 50% of the issue amount. Then, we eliminate 2,425 trade reports in which the volume was reported as less than $1,000.

Remaining unmatched trade reports

After the steps to assemble the data set, there are still unmatched trade reports. Recall that every agency chain should generate at least two agency trade reports and every principal-to-principal should generate two dealer trade reports or a single dealer-customer trade report. This implies that if we have cleaned our dataset perfectly, there should be no unmatched agency or interdealer principal-to-principal trade reports. In fact, we are left with 835,426 unmatched agency trade reports, of which 429,578 are sells and 405,848 are buys, and 1,360,771 unmatched interdealer principal-to-principal trades reports, of which 585,407 are sells and 775,364 are buys. (These numbers are not shown in Table B1).

In addition, when we examine when these unmatched trade reports occurred, we find that they occur more frequently at the beginning of the sample period. For example, in August 2002, unmatched agency trade reports are 47% of all agency trade reports, while by August 2006 (near the end of our sample), it is 13%. Likewise, in August 2002, unmatched dealer trades are 27% of all trades, while by August 2006, it is close to 3%.

There are several possible reasons why unmatched reports remain in our data.

1. An unmatched report may represent a trade that occurred, but the other dealer failed to submit her own report.
2. An unmatched report may represent a trade that was modified, cancelled, or reversed, but we cannot identify it as such. For example, Table B1 shows there are 30,724 trade reports marked as reversals that were never matched to an earlier trade.
3. There may be two unmatched reports that describe the same trade. However, because of differences in reporting conventions or mistakes in data entry, the reports do not match.

If we knew why each an unmatched trade report exists, then we could process it appropriately. For example, if a trade report were of type 1, we would like to keep it in our sample. If a trade report were of type 2, we would like to eliminate it from our sample. If there is a group of trade reports of type 3, then one solution would keep all the buy trade reports or keep all the sell trade reports. Unfortunately, we do not know how to distinguish among different types of unmatched trade reports.

Therefore, to check that our results are robust to the various types of possible errors, we construct three different final trade sets: 1) keeping all unmatched agency and interdealer principal-to-principal trades (both buy and sell), 2) dropping all unmatched agency and interdealer principal-to-principal trades (both buy and sell), and 3) we drop all unmatched buy reports in interdealer principal-to-principal trades, drop trade reports that are replicates of agency chains and principal-to-principal trades, and unmatched dealer sells where the difference between execution date and reporting date is greater than 1 day (more than 50% of these trades where 1 day). Table B1 shows that we have dropped 2,196,197 unmatched trade reports between the first sample and third sample.

In the paper, we report estimates using the conservative sample, but we have re-run all estimates in Table 3-11 using the moderate and aggressive samples as well. The results are similar across these three samples and are available upon request.

Constructing the Cleaned Phases Sample

The entire dataset of cleaned bonds is not necessarily useful when evaluating the effect of TRACE. Our empirical strategy is based on comparing a bond's trading behavior when it changes from non-disseminated to disseminated. If a bond does not have both disseminated and non-disseminated trades, it will not be useful for many of our statistical tests. Many bonds are disseminated for their entire trading history. These include bonds that belong to a FINRA Phase that are issued after the beginning of the Phase date, and bonds that may be issued before a Phase begins but only trade after the dissemination change date for that Phase. There are also bonds that are always non-disseminated. These are bonds that may mature before the beginning of their Phase date as well as bonds that belong to a Phase but never trade after the Phase begins.

Table B2 outlines the steps from Academic Cleaned Sample in Table B1 to the Cleaned Phase Sample, the sample of bonds which exist and have zero or non-disseminated trading before the start of a Phase and zero or disseminated trading after the start of a Phase. We begin with a list of all Phase 1, 2, 3A, and 3B bonds. There are 20,713 bonds in this list, of which 20,337 exist in the Cleaned Academic

TRACE Sample. They have 18,390,345 trades during our sample period. Thus, about two-thirds of the bonds in our Cleaned Academic TRACE Sample are in our Phase list and this represents 82.4% of the trades.

For Phases 2, 3A, and 3B, FINRA provided us with a list of bonds that began being disseminated at the start of each of the three Phases. This list was provided in a non-electronic format where bonds were identified with ticker symbols. Unfortunately, many ticker symbols longer than six characters were truncated. This was a problem for firms with a four-character company ticker symbol which also issued bonds with three-character security tickers. In particular, many GMAC bonds were truncated. Since FINRA also provided us with coupon and maturity dates for each bond, we were able to hand-match many of the truncated ticker symbols, but not all. The list of Phase 1 bonds was not provided by FINRA, and we generated it ourselves. If a bond has disseminated trades before the start of Phase 2, and is not part of the Phase 2, 3A, 3B samples or the FINRA50/120 samples, we classify it as a Phase 1 bond.

In addition to the four Phases that correspond to the FINRA dissemination dates, FINRA also maintained two other lists of bonds, which we call the FINRA50 and the FINRA120.[5] The FINRA50 represent 50 Non-Investment Grade (High-Yield) securities disseminated under Fixed Income Pricing System (FIPS2). This list of 50 bonds changes over time with bonds both entering and exiting. FINRA lists all of these bonds on their website and there are a total of 149 bonds that were in the FINRA50 at some point during its existence from July 1, 2002 until July 14, 2004. The FINRA120 list is a special set of 120 investment grade rated Baa/BBB that FINRA delayed Phase 2 dissemination for. Phase 2 dissemination started on March 3, 2003 for Phase 2 bonds but started on April 14, 2003 for the FINRA120. This special sample was created so that FINRA could conduct a controlled experiment to study the effects of dissemination in Phase 2, contained in Goldstein, Hotchkiss, and Sirri (2007).

Table B2 explains how we went from FINRA's list of Phase 2, 3A, and 3B bonds, and our list of Phase 1 bonds, to our Cleaned Phase Sample. For each Phase list, we only use bonds that exist in our Cleaned Academic TRACE Sample. Some bonds on the FINRA lists did not trade during our sample period and thus are not in the Cleaned Academic TRACE Sample. This is shown between lines 1 and 2 under Phases 2, 3A, and 3B.

We next eliminate any bonds that also exist in the FINRA50 or FINRA120 list. Following this, we eliminate bonds that do not exist (i.e., were not issued or matured) during the period 90 days before until 90 days after the start of the Phase. Finally, we dropped some bonds with data problems. There are a few bonds where FINRA report disseminated trades before the start of the Phase, or non-disseminated trades after start of Phase. After applying these steps for each Phase list, what remains is our cleaned sample by Phases. There is a total of 16,854 bonds in our Cleaned Phase Sample representing 15,952,736 trades during our time period.

[5] The list is available at http://www.finra.org/Industry/Compliance/MarketTransparency/TRACE/Announcements/P117685, last accessed March 7, 2019.

Appendix C: NAIC Data

The National Association of Insurance Companies (NAIC) dataset we use is from Mergent FISD available on WRDS. The NAIC requires insurance companies to self-report all securities transactions in their financial statements. There are 63,859 bonds with 1,933,095 reported transactions in the NAIC file over the period January 1, 2000 until December 31, 2006. Schedule D of the annual NAIC filings require insurers to report all bond transactions in one of three categories: bonds added to the portfolio during the calendar year and held through the end of the year, bonds deleted from the portfolio during the calendar year that were not added in the same year, and bonds added and deleted in the same calendar year. For each transaction, the database records the CUSIP, date, par value of the transaction, the actual value of the transaction, if it was an addition or deletion, and a field for the counterparty involved in the transaction. Prices are not reported but can be computed from the ratio of the value received in the transaction to the par value of the bonds in the transaction. Importantly, the names of the insurance companies involved in the transactions are excluded from the data.

To make the NAIC analysis comparable to the TRACE analysis, we first match our sample of NAIC bonds with the Academic TRACE sample by CUSIP. The universe of bonds which insurance companies trade is much larger than that reported by TRACE. 53,030 NAIC bonds representing 1,150,611 reported transactions are not included in our Academic TRACE sample and are eliminated. Table B1 reports the number of transactions and CUSIPs eliminated by this step.

Next, we eliminate reported transactions that are not connected to trades. The NAIC database contains all transactions involving insurance companies' bond portfolios, not only buy and sell transactions, but also other transactions such as bond calls and maturities. The type of transaction is coded in the counterparty field. We eliminate all transactions that change bond portfolio holdings that are not buys or sells. These include the following codes: CALL, CANCEL, CONVERT, EXCHANGE, ISSUE, MATURE, PUT, REDEEM, SINKING FUND, TAX-FREE EXCHANGE, TENDER, TRANSFER, PAYDOWN, and REPLACE.

There are two prevalent entries in the counterparty field comprising almost 15% of the cleaned database that required additional attention: DIRECT and VARIOUS. DIRECT may indicate a direct placement, similar to an underwriting, or it may indicate the name of a counterparty in an actual trade. VARIOUS is simply an ambiguous catch-all, where some records may be actual trades and some are not. To check whether DIRECT and VARIOUS represent actual trades, we matched these NAIC records to TRACE using the CUSIP, price, volume, and date of the transaction. For DIRECT, only about 3% of transactions match into the TRACE dataset, while for VARIOUS only about 1% of transactions match. Because of the problems identifying which of the DIRECT or VARIOUS transactions are actual trades, we eliminate them along with the other codes listed above that are not buys and sells. As shown in Table C1, all such filters eliminate 200,639 reported transactions on 8,588 different bonds.

We eliminate a small number of trades with data issues, i.e. missing prices, negative prices, etc. We next eliminate trades with timing issues, i.e., trades that are executed before or on the bonds' offering or after or on the bonds' maturity date. Finally, a large fraction of NAIC transactions take place

on the offering and maturity dates. We believe that this is because insurance companies are large customers of bond offerings and hold many bonds until maturity. The NAIC rules require its members to list purchases as a transaction since the bonds are added to their portfolio. Since these transactions are probably part of the underwriting, we do not include them as trades. If an insurance company holds the bond until its maturity, that transaction will also be recorded by NAIC but we eliminate it. Finally, we also exclude transactions listed on bond holidays. These screens shown in Table C1 are similar to those applied to the TRACE dataset in Table B1.

After the screens and matching, there are 9,997 bonds and 505,753 reported transactions (which we believe to be buys and sells) in our "clean" NAIC sample. Importantly, the NAIC time period in Table C1 is thirty months longer than the TRACE time period in Table B1. When we restrict to the time period July 1, 2002 until December 31, 2006, there are 9,379 bonds and 332,224 transactions, as shown in the last three columns of Table C1.

As mentioned in Appendix A, we believe that many trades in TRACE are disaggregated by the NAIC reporting process. When comparing the NAIC and TRACE databases, there are multiple NAIC transactions that match to a single TRACE trade using CUSIP, date, price and counterparty, but not volume. However, if we group NAIC transactions by CUSIP, date, price and counterparty into a single record with a combined volume, many of these grouped NAIC trades match to a corresponding single trade in TRACE.

There are two reasons that trades are disaggregated in NAIC. The first reason is how NAIC requires transactions to be reported on Schedule D of the annual NAIC filing. Insurers must separately report bonds purchased and sold in the same year from bonds purchased and held through the end of the year. This means if an insurance company purchases $1 million par of a bond on January 1, 2001 and sells $500,000 of this before December 31, 2001 and the remaining $500,000 sometime in the following year, under NAIC reporting instructions, this single purchase would be split into two separate purchases of $500,000 each, reported in two different sections of Schedule D. One $500,000 purchase would be reported in the long-term purchase reporting section, and one $500,000 purchase would be reported in the short-term holding section.

When the NAIC database is compiled, the above trade would appear as two purchases of the same bond occurring on the same day at the same price. In TRACE, however, the dealer who sold the bond would report this as one $1 million trade. If we aggregate the volume of the NAIC trades that occur in the same bond, on the same day, at the same price, the NAIC transaction would match to the TRACE trade, as a single trade. It's worth noting that since the insurance company sold the bond holdings as two separate pieces of $500,000 each on two separate days, two distinct sales of $500,000 would be reported as two sales in both NAIC and TRACE.

A second reason for why a single trade may be reported as multiple trades is that distinct subsidiaries of an insurance company may book portions of a trade to their respective division, and each division makes its own statutory filings to the NAIC. This can occur, for instance, if part of a trade is

allocated to the property and casualty group and another portion allocated to the life insurance group. In the NAIC database, this appears as two trades, while in TRACE, it appears as one trade.

We attempt to correct for these two reporting problems by grouping transactions that we believe correspond to the same trade. Any records that share the same date, CUSIP, counterparty, transaction type (buy or sell), and have prices within 1 cent of another are grouped and considered a single trade. We show this grouping in Table C1. In the cleaned NAIC file, from January 1, 2000 to December 31, 2006, the number of trades reduces from 505,753 to 412,758.

Table C1 reports the steps we took to process the raw NAIC file into our cleaned NAIC database. We only use those bonds from the NAIC database that are also in the Cleaned Academic TRACE database for our analysis.[6] Because of the misreporting issue discussed above, Table C1 reports the total number of transactions from the NAIC database in the column labeled "Ungrouped Trades." It also reports an estimate of the true number of trades by grouping transactions with identical CUSIP, date, price, and counterparty into a single record with volume summed for the grouping. These are labeled "Grouped Trades" in a separate column in Table C1. From July 1, 2002 to December 31, 2006, the clean NAIC database contains 9,379 bonds. There are 332,224 ungrouped trades, which correspond to 272,133 grouped trades. This compares to 22,582,689 trades on 30,814 bonds in the Cleaned Academic TRACE database over the same period.

To assign the bonds in NAIC to a TRACE Phase, we simply match the cleaned Phase list from TRACE used in Table B2 to the sample of cleaned NAIC bonds. Table A2 reports the number of NAIC CUSIPs, and both grouped and ungrouped trades in each Phase. Importantly, in Phase 1, we match 366 CUSIPs out of 388 TRACE Phase 1 CUSIPs.

As discussed in Appendix A, grouping trades does not affect our NAIC volume analysis. However, the price standard deviation increases when we group trades. We, therefore, report the analysis of NAIC trades both with and without grouping.

[6] 45,902 bonds in the NAIC database are not in the Cleaned Academic TRACE database. A large fraction of these bonds are SEC Rule 144a bonds. SEC Rule 144A bonds are not covered by TRACE during our sample period.

Table B1. Steps from Academic TRACE File to Cleaned Academic TRACE Sample

	Affected CUSIPs	Affected trade reports	Remaining trade reports
	(1)	(2)	(3)
Source: Academic TRACE File	41,676	37,934,432	
Eliminate bonds based on characteristics			
Bonds unmatched to FISD by CUSIP	2,178	365,504	37,568,928
Convertible bonds	2,154	2,655,202	34,913,726
Exchangeable bonds	129	43,688	34,870,038
Other equity-linked bonds	625	210,005	34,660,033
SEC Rule 144a bonds	4,587	1,287,935	33,372,098
Bonds with missing issue size or issue size < 100,000	345	89,727	33,282,371
Eliminate trades because of self-reported errors			
Modifies: Trade reports matched to earlier report using sequence number	20,732	1,250,351	32,032,020
Modifies: Trade reports not matched to earlier report	0	0	32,032,020
Cancels: Trade reports matched to earlier report using sequence number	22,353	1,013,223	31,018,797
Cancels: Trade reports not matched to earlier report	0	0	31,018,797
Reversals: Trade reports matched to earlier report using nine-way match	21,278	841,108	30,177,689
Reversals: Trade reports matched to earlier report using eight-way match	8,725	87,250	30,090,439
Reversals: Trade reports matched to earlier report using eight-way match (with price rounding to two decimals)	1,137	5,694	30,084,745
Reversals: Trade reports not matched to earlier report	7,440	30,724	30,054,021
Modifying trade reports based on TRACE submission rules			
Add in extra trade reports for Locked-in trades	18,446	523,125	30,577,146
Eliminate trade reports with timing issues			
Trade reports executed earlier than July 1, 2002	956	2,195	30,574,951
Trade reports executed before bond's offering date (i.e., when-issued market)	6,770	291,840	30,283,111
Trade reports executed after bond's maturity date (i.e., bankrupt bonds)	347	93,287	30,189,824
Trade reports that occur on SIFMA holidays	20,775	802,065	29,387,759
Eliminate redundant trade reports			
First pass: Using Original IDs			
Agency chains, exact volume, original buyer-seller IDs	20,486	1,290,607	28,097,152
Agency chains, aggregate volume, original buyer-seller IDs	4,131	57,314	28,039,838
Agency chains, exact volume, original buyer-seller IDs, misreported agency/principal	13,139	176,744	27,863,094
Agency chains, aggregate volume, original buyer-seller IDs, misreported agency/principal	1,008	3,736	27,859,358
Agency chains, exact volume, misreported original contraparty ID	7,725	94,587	27,764,771
Agency chains, aggregate volume, misreported original contraparty ID	1,398	11,241	27,753,530
Interdealer principal-to-principal trade, exact volume, original buyer-seller IDs	26,350	3,713,491	24,040,039
Interdealer principal-to-principal trade, aggregate volume, original buyer-seller IDs	4,509	56,605	23,983,434
Interdealer principal-to-principal trade, exact volume, original buyer-seller IDs, misreported agency trade report	17,430	508,709	23,474,725
Interdealer principal-to-principal trade, aggregate volume, original buyer-seller IDs, misreported agency trade report	1,142	4,867	23,469,858
Interdealer principal-to-principal trade, exact volume, misreported original contraparty ID	17,002	353,622	23,116,236
Interdealer principal-to-principal trade, aggreage volume, misreported original contraparty ID	1,734	10,733	23,105,503

Second pass: Substitute in common dealer ID pairs from misreported original counterparty IDs			
Agency chains with all modifications using substitute IDs	965	2,963	23,102,540
Interdealer principal-to-principal trade with all modifications using substitute IDs	4,146	18,225	23,084,315
Third pass: rounding prices			
Agency chains with all modifications using rounded prices	6,079	246,541	22,837,774
Interdealer principal-to-principal trade with all modifications using rounded prices	15,306	213,864	22,623,910
Fourth pass: truncated prices			
Agency chains with all modifications using truncated prices	1,360	10,117	22,613,793
Interdealer principal-to-principal trade with all modifications using truncated prices	5,108	16,152	22,597,641
Modify or eliminate trade reports with price or volume issues			
Modify prices in trade reports that are vastly out of line	2,569	6,850	22,597,641
Eliminate trade reports with price less than 0.01 percentile and greater than 99.99 percent	284	4,848	22,592,793
Eliminate trade reports with volume less than 0.01 percentile and greater than 99.99 perc	1,564	3,283	22,589,510
Eliminate trade reports with volume/issue amount ≥ 50%	1,868	4,396	22,585,114
Eliminate trade reports with volume less than 1000	693	2,425	22,582,689
Cleaned Academic TRACE Sample keeping all unmatched trade reports	**30,814**	**22,582,689**	
Unmatched dealer buy trade reports	24,916	968,799	21,613,890
Unmatched dealer sell trade reports	21,904	783,012	20,830,878
Unmatched agency trade reports	15,451	444,386	20,386,492
Cleaned Academic TRACE Sample keeping unmatched dealer sell trade reports	**30,728**	**21,613,890**	
			2,196,197
Cleaned Academic TRACE Sample dropping all unmatched trade reports	**30,676**	**20,386,492**	

Notes. Filters are applied sequentially. This table reports the steps from the historical TRACE file to the Clean Historical TRACE file. Other equity-linked bonds have "KNOCK", "REVERSE", "EQUITY", "LINKED", and "TBD" in the bond's FISD issue name. A seven-way match is based on CUSIP, execution date, execution time, price, quantity, buy-sell indicator, and dealer-customer indicator. A six-way match drops the execution time requirement. SIFMA holidays correspond to "Recommended Early Close" and "Recommended Full Close" dates listed at http://www.sifma.org/uploadedfiles/research/statistics/statisticsfiles/misc-us-historical-holiday-market-recommendations-sifma.pdf. In the third pass, prices are rouned to two decimal places. In the fourth pass, prices are truncated to two decimal places. We use the rounded price for all subsequent steps after the fourth pass. We define a price of a trade record to be vastly out of line if the magnitude of difference between the price of that trade record and the volume weighted median of 10 records before and 10 records after is greater than 20 per $100 par and sign of difference before and after are opposite. We require that all the 10 trades to be within the 30 days before and 30 days after. If there are fewer than 10 trades in the 30-day window,

Table B2. Steps from FINRA's Phase Listings to Cleaned Phase Sample

	CUSIPs (1)	Trade records (2)
Cleaned Academic TRACE Sample (after Table B1) keeping all unmatched trade reports	30,814	22,582,689
Source: FINRA list of Phase 1-3B bonds	20,713	-
Bonds on both FINRA Phase list and Cleaned Academic TRACE Sample	20,337	18,390,345
Phase 1		
list of Phase 1 bonds*	464	4,910,313
bonds in FINRA50 at start of phase	3	33,172
bonds do not exist as of start of phase	65	745,749
bonds do not exist for the entire period 90 days before until 90 days after start of phase	8	536
Cleaned Phase 1 Sample	388	4,130,856
Phase 2		
FINRA's list of Phase 2 bonds	3,203	-
Phase 2 bonds in Cleaned Academic TRACE Sample	3,016	3,111,987
bonds in FINRA50 before or at start of phase	0	0
bonds do not exist as of start of phase	265	22,283
bonds do not exist for the entire period 90 days before until 90 days after start of phase	225	155,075
Cleaned Phase 2 Sample	2,526	2,934,629
Phase 3A		
FINRA's list of Phase 3A bonds	13,364	-
Phase 3A bonds in Cleaned Academic TRACE Sample	13,197	8,855,575
bonds in FINRA50 or FINRA120 before or at start of phase	78	651,859
bonds do not exist as of start of phase	969	181,008
bonds do not exist for the entire period 90 days before until 90 days after start of phase	1,069	280,214
Cleaned Phase 3A Sample	11,081	7,742,494
Phase 3B		
FINRA's list of Phase 3B bonds	3,682	-
Phase 3B bonds in Cleaned Academic TRACE Sample	3,660	1,512,470
bonds in FINRA50 or FINRA120 before or at start of phase	26	58,881
bonds do not exist as of start of phase	643	257,195
bonds do not exist during the period 90 days before until 90 days after start of phase	132	51,637
Cleaned Phase 3B Sample	2,859	1,144,757
Total Cleaned Phase 1-3B Sample	**16,854**	**15,952,736**

Notes. This table reports the match between the Cleaned Academic TRACE file and FINRA's Phase Listings. Not all bonds in the TRACE Academic Sample are classified in a FINRA Phase. Excluded bonds are those issued after 7/1/02 that are always disseminated and those that mature before 2/7/05 that are never disseminated. We construct the Phase 1 list by including all bonds with disseminated trades before Phase 2 that are not on the FINRA Phase 2, 3A, or 3B lists. The Phase 2, 3A, and 3B lists were obtained directly from FINRA. The FINRA50 and FINRA120 lists are from www.finra.org. Bonds in FINRA's Phase lists that are not in the Cleaned Academic TRACE Sample have either never traded during the sample period or have been eliminated due to cleaning process in Table A1.

Table C1. Steps from Historical NAIC File to Cleaned Academic Sample

	January 1, 2000 - December 31, 2006			July 1, 2002 - December 31, 2006		
	CUSIPs	Ungrouped trades	Grouped trades	CUSIPs	Ungrouped trades	Grouped trades
Original Source: NAIC Transactions File	63,859	1,933,095	1,490,831	50,968	1,341,769	1,032,125
Match NAIC Bonds with Cleaned Academic TRACE Sample keeping all unmatched trade records						
CUSIP not found in Cleaned Academic TRACE Sample keeping all unmatched trade records	53,030	1,150,611	884,359	40,454	819,448	629,630
Eliminate transactions which are not trades						
Non-trade indicated by counterparty field entry (calls, converts, etc.)	8,588	200,639	145,052	8,124	155,205	106,278
Eliminated trades with timing issues						
Trades executed on or before bond's offering date	3,813	53,580	29,053	2,000	20,474	11,856
Trades executed on or after bond's maturity date	556	1,026	944	556	1,026	944
Trades executed on weekend or SIFMA Holiday	5,086	20,131	17,661	3,789	11,963	10,449
Post July 2002 trades executed on days with no TRACE trades**	4	8	7	4	8	7
Eliminate trades with price and volume issues						
Trades with price less than 0.01 percentile and greater than 99.99 percentile	276	564	376	271	559	371
Trades with volume/issue amount >= 50%	202	472	318	126	294	191
Trades with volume less than 1000	176	311	303	158	270	265
Cleaned NAIC Sample	9,997	505,753	412,758	9,379	332,224	272,133

Notes: Filters are applied sequentially. The CUSIPs column gives total number of CUSIPs eliminated from the database by adding that row's filter. The trades column gives total
number of observations eliminated by adding that row's filter.**On June 11, 2004, the SEC declared a holiday when President Reagan died.
Grouping is done if the difference in Price is <= |0.01| and the day, counterparty, insurer type, and buy or sell are equal.