Online Appendix to

"Middlemen Matter: Corporate Bond Market Liquidity and Dealer Inventory Funding"

Andreas C. Rapp

Tilburg University - Department of Finance

This version: October 2017

This online appendix contains additional information to accompany the paper "*Middlemen Matter: Corporate Bond Market Liquidity and Dealer Inventory Funding*". The contents are as follows:

- **A.1** (Data Appendix) details the steps required to process the raw data.
- **A.2** (Model Appendix) derives the the price impact regression model.

A.1 Data Appendix

In this Section I lay out the steps required to process the raw data where I closely follow the procedure used by Asquith et al. (2013). Tables A-1 and A-2 outline the effects of the steps on the number of bonds and trades. The sample period stretches over 12 years and ranges from July 1, 2002 to June 30, 2014.

Step 1: Cleaning NAIC transactions

The NAIC requires insurance companies to self-report security transactions in their financial statements. Each entry contains detailed transaction information including the CUSIP, the date of the transaction, the par value of the transaction, the actual value received in the transaction, the (clean) price on a $100-par-basis, the buy/sell indicator, and a field for the counterparty involved in the transaction.[1]

Screening on bond characteristics: In the screening among corporate bonds I essentially stick to the literature. I drop all non-corporate bonds using issuers' industry codes and issue bond-type references.[2] Furthermore, I discard bonds with convertibility and exchangeability features,[3] bonds that are putable, bonds issued in a foreign currency, bonds that fall under the SEC 144a rule for a private placement (i.e., these issues do not require TRACE reporting), and finally all forms of asset-backed securities or secured lease obligation issues. Now, I am left with a sample of only *corporate bonds*. Here, I keep all coupon types (with the exception of perpetual bonds) independent of their security level. Lastly, I hold on to callable bonds because they comprise a substantial part of the remaining sample (i.e., 38%). As in Bao et al. (2011) my analysis has an exclusive focus on investment-grade bonds.[4] For one, this ensures the assessment and classification of my results in a comparable context. For another, due to regulatory constraints insurance companies are not a natural player in high-yield bond markets. In fact, their (secondary-market) trades in non-investment grade

[1]The identification of counterparties is one-sided. That is, the names of the insurance companies involved in the transactions, as available to Ellul et al. (2011) and O'Hara et al. (2016), are not given in my dataset.

[2]This includes industry codes for *Government*, *Unassigned*, and *Miscellaneous*. As insurance companies are common takers of government bonds this step reduces the NAIC bond universe by nearly half. Using bond type references I drop the remaining U.S. agency debt, foreign government bonds, and municipality bonds.

[3]This includes already converted bonds, which are listed under preferred equity.

[4]As with covariance-based liquidity measures, arguably though to a lesser extent, my regression-based methodology to bid-ask spreads improves with more observations (i.e., more frequent trading per bond). High-yield bonds are traded even less frequently than regular bonds such that parameter estimates obtained separately for this bond segment are potentially noisy and unreliable.

bonds are oftentimes the result of regulatory induced sell-offs and thus are not representative for a typical high-yield trade (Ellul et al., 2011). The final round of filters apply to records with potential data issues concerning their price (missing, negative, or unreasonably large prices), volume (deletion of non-institutional trades < \$100,000; trade sizes > 50% of the issuance amount are dropped; observations > \$20 million (98.5^{th} percentile) are truncated), or timing (trades on a bond's offering or maturity date, and trades on SIFMA or SEC declared trading holidays).

Separation of counterparties: The counterparty field is key to dealer identification. As reports are manually coded it needs extensive cleaning with respect to name variations and misspellings though. With approximate string matching algorithms I create a dictionary that contains close to 41.000 unique entries and links to 764 counterparty keys.[5] A counterparty key captures insurers' trading partners in a broad sense. I differentiate between actual dealers (where I bundle all trading desks within a particular dealer firm), other trading partners (e.g., other insurance companies), and portfolio related events that do not represent buys and sells (such as bond calls, maturities, conversions, transfers, redemptions etc.).[6] Moreover, there are two entries – DIRECT (0.8%) and VARIOUS (12%) – that cannot be linked to dealers or "bond events" and make up approximately 13% of all cleaned NAIC entries. DIRECT likely refers to a direct placement of bond as opposed to a public offering. And VARIOUS is probably a generic catch-all fill-in that points to records that may be actual trades with dealers but may well be other no-trade related portfolio transactions.[7] In fact, there is no clear reason why insurers used VARIOUS instead of a dealer or a "bond event". Following Asquith et al. (2013) I also check whether DIRECT and VARIOUS represent actual TRACE trades and include them in my matching approach. The matching success is far below average and amounts to less than 2% of total volume in the cleaned Enhanced TRACE dataset. As I can neither distinguish whether DIRECT or VARIOUS resemble actual trades nor link the keys to a CDS spread I eliminate these entries. Similarly, I drop all transactions that are grouped in the LEFTOVER category. Then, excluding non-dealers and all remaining "bond events", and linking dealer subsidiaries

[5]In comparison to Asquith et al. (2013) who group 11% of transactions in a LEFTOVER category my dictionary largely improves the grouping of counterparty entries. Trades that do fall in a LEFTOVER category (where I distinguish between UNIDENTIFIABLE (1.8%) and UNASSIGNED (1.3%) transactions) only represent on average close to 3% of transactions per year of data. These are entries that are outright incomprehensible, or appear infrequently and cannot be clearly assigned to a key.

[6]Here, MATURITY (7.9%) accounts for the biggest group of "bond events'; other categories are CALL (2.7%), TAX(1.8%), TRANSFER (1.3%), REDEMPTION (0.9%), TENDER (0.8%), EXCHANGE (0.6%).

[7]When matching their NAIC sample with TRACE trades, Asquith et al. (2013) find that DIRECT comprises about 3% of transactions and VARIOUS only about 1% of transactions.

to their parent companies I end up with 514 unique dealer firms over the sample period of 12 years. The increase in unique counterparty names in comparison to Asquith et al. (2013), who find 105 (86 consolidated) dealers, is likely "long-tail" effect. That is, while the bulk of transaction have to be attributed to a small number of firms (the most active 25 dealers account for a bit more than 80% of all transactions) there exists a large number of dealers with relatively small number of transactions.

Step 2: Cleaning Enhanced TRACE transactions

The FINRA's TRACE database is the result of regulatory initiatives to increase transparency in corporate bond markets. The FINRA operates the reporting and dissemination facility for corporate bond trades. It is responsible for making transaction information available to market participants since July 1, 2002. The transparency initiative was introduced in three phases (Phases 1, 2, 3A, and 3B) where each time the reporting requirements and the number of bonds that required reporting were expanded. Since February 7, 2005 close to all corporate bond transactions need to be reported.

Dealers who are FINRA members must self-report their bond trades to the regulator immediately. Dissemination of transaction data to the public happens with a 15-minute lag. TRACE reports actual trade sizes for par values of $5 million or smaller for trades in investment grade bonds. All larger trades are capped (i.e., marked with "5MM+"). Since March 2010 FINRA releases an *Enhanced* TRACE dataset, which includes both disseminated and non-disseminated transaction records and captures full trade sizes, starting from TRACE's initiation in July 2002. The data is available with an 18-month lag.[8] In this paper, I use the *Enhanced* TRACE database where a trade report includes the CUSIP, the trade's execution date and time, the clean price on a $100-par-basis, the volume traded (in $ of par), a buy/sell indicator, and whether the trade was a customer-dealer or an inter-dealer trade.

Since April 2017 TRACE datasets with anonymized dealer IDs are available for purchase from FINRA. Importantly, reverse-engineering dealer identities using these IDs is contractually prohibited. In comparison, these datasets allow a clear differentiation (but not identification) of dealer trading desks' trade flows through anonymized dealer IDs. My TRACE trade data is anonymous with respect to dealer identities. In this paper dealer identification is achieved through matching of TRACE and NAIC data.

[8]One can therefore differentiate between the *Public* data which is available within 15 minutes and the 18-month lagged *Enhanced* data.

Eliminating erroneous entries: Most of the cleaning involves eliminating erroneous trade reports (Dick-Nielsen, 2009, 2014), e.g., cancellations, modifications, reversals, or duplicates. The NAIC transactions are customer-dealer trades by definition and thus for matching I discard all TRACE inter-dealer trades. Furthermore, I eliminate all agency trades where dealers essentially act as brokers and do not build up inventories. Among principal trades I adjust trade prices for potential "markups or markdowns" (e.g., potential commissions paid) whenever the buy or sell commission field is non-empty. Lastly, I drop trade reports on a bond's issuance date, all non-secondary market trades, trades with a special price flag or an irregular trade type, as well as trades connected to equity-linked notes.

Screening on bond characteristics: Since the retrieved TRACE data is based on bonds from my cleaned NAIC dataset they are subject to the same FISD filters. Similarly, I eliminate all trade reports with potential data errors concerning the price, volume, or timing of the transaction.

Step 3: Matching NAIC and TRACE

The key innovation of my dataset is the ability to link dealer identities with transaction prices and allow for an empirical identification at the individual transaction-level. This is achieved by matching the transactions of the cleaned NAIC dataset with those in the cleaned Enhanced TRACE dataset. Specifically, I use five trade criteria to form a match: the CUSIP, the trade execution date, the trading volume, the buy/sell indicator, and the price. The matching is exact on the first four criteria and approximate on the last where I allow for a price difference of one or less than one cent (i.e., $\leq |0.01|$ on a \$100-par-basis).[9] The exact match on volume seems most appropriate given that NAIC trade prices are based on the ratio of the value received to the par value of a bond in each transaction.

As pointed out by Asquith et al. (2013), the NAIC database may suffer from a systematic error due to a disaggregation of trades given their reporting process that leads to an over-reporting in the number of trades and an under-reporting of the true price dispersion.[10] Specifically, multiple NAIC transactions match a single TRACE trade using CUSIP, trade execution date, price, and counterparty but taken separately show a discrepancy with the

[9] Both NAIC and TRACE transactions are self-reported and price discrepancies due to entry errors or rounding are possible. This may lead to (small) differences in the reported prices.

[10] Reasons for this are: First, the reporting requirements demand that changes to a portfolio position, depending on the length of the holding period, need to be filed as if they constitute separate trades starting with the opening of the initial position. Second, subsidiaries of an insurance company, handing in their own statutory filings to the NAIC, may book and report portions of one trade to their respective division.

reported volume. After grouping (on bond CUSIP, trade execution date, trade counterparty, buy/sell indicator, and insurer type) and aggregating NAIC volumes these aggregate trades give a better match to a corresponding single TRACE trade. Thus, in order to correct for the disaggregation I match NAIC and TRACE trades in three rounds: In a first round, I match *non-aggregated* (i.e., as reported) NAIC trades with TRACE trades. In a second round, I group NAIC trades per bond CUSIP when trade execution date, counterparty, and buy/sell indicator are identical and the difference in price is smaller or equal to one cent (i.e., $\leq |0.01|$ on a \$100-par-basis). Only then I match these *aggregated* NAIC trades with the TRACE trades.[11] In the last round I additionally group NAIC trades on the variable *insurer type* before I complete the matching. That is, the second round accounts for the possibility that an insurance company may book and report portions of the same trade under their respective divisions or subsidiaries.

Essentially building the sample only on TRACE trades reduces potential issues due to various data entry errors in NAIC (e.g., excessive price changes, disaggregation of trades and consequently more stale prices) and should clearly decrease the statistical noise in the price impact regression. The matching with TRACE trades serves two purposes: (i) to identify a dealer behind a bond transaction and subsequently links dealer characteristics to a respective transaction price; (ii) to keep a transaction time stamp in order to have a time ordering of trades. The latter point is important too: since the econometric model in equation (A.2.14) relies on the assumption that trades are appropriately ordered in time within a trading day I gain statistical power over a sample that lacks transaction times (see Bessembinder et al. (2006)).

The average matching success is at 42.1% percent per year.[12] I refer to a *matched* TRACE trade in case the dealer identity behind the particular trade report is known whereas I refer to an *unmatched* TRACE trade in case the dealer identity is unknown. In total I am left with 295,424 matched TRACE trades involving 410 different dealers and 12,059 bonds of 2,309 issuers.

Step 4: Mapping Dealer CDS Spreads

To link dealers with a CDS spread I bundle trading desk within a dealer firm and then

[11] Aggregation takes place before trades smaller than \$100,000 are dropped.

[12] Matching success is a function of the permitted deviation in the approximate price match. As I become less conservative (e.g., allowing for a price difference of more than one cent) the matching success increases. The fact that a large fraction of NAIC transactions still cannot be matched with TRACE data is a concerning issue that could well be directed to the NAIC.

determine its relevant parent company. This is so for two reasons: First, CDS contracts are usually not written on subsidiaries or business units but only the parent company. Second, some of the dealers captured in the NAIC trade reports are listed under separate counterparty names but are really part of the same entity at the time of the trade.[13] In doing so, I take merger and acquisition activities among dealer firms into account.[14] Using SDC Platinum and Zephyr, I account for 234 changes in dealers' parent companies due to mergers and acquisitions, or bankruptcy. As a dealer's parent company changes I combine the trading activity under the successor's name.

I collect CDS spreads on senior unsecured debt with a five-year maturity from Bloomberg and Datastream/CMA respectively.[15]

Using CDS spreads comes with three limitations: First, overall CDS coverage is not complete. While CDS contracts are available for bigger institutions there are often no contracts for smaller non-bank dealer boutiques.[16] Second, I do not have access to all data providers (e.g., the Markit database). As a result, I am missing seven of the most active 25 dealers in my sample of matched TRACE trades.[17] Overall, I collect CDS spreads on 69 dealers (in terms of matched TRACE trades this resembles 100% of the 10 most active dealers, and 87% (75%) of the 25 most active (all) dealers). Third, some series only start after July 1, 2002, end before June 30, 2014, have gaps, or show periods of stale prices (i.e., if a spread does not update for more than two weeks I treat it as missing).[18]

To retain the widest possible cross-sectional coverage with respect to dealer-specific inventory costs, I fall back on long-term credit ratings in case I do not have a dealer's CDS spread. Rating histories are obtained from Moody's. This way I still capture prominent non-bank dealers active in the U.S. corporate bond market. If available I collected rating data on the dealers that make up 99% of all matched TRACE trades

[13]Joint entities that make up for a lot of trades are, e.g., Salomon Brothers (acquired by Travelers Group (now Citigroup) in 1998), Spear Leeds & Kellogg (belongs to Goldman Sachs since 2000), or Donaldson Lufkin & Jenrette (part of Credit Suisse since 2002).

[14]For example, big mergers in terms of transactions are, e.g., Bear Stearns and J.P. Morgan Chase as well as Wachovia and Wells Fargo both in 2008.

[15]My access to CMA is limited to series after January 2004 and before September 2010. Bloomberg, on the other hand, offers CDS data for most big institutions starting with TRACE initiation in July 2002.

[16]For example, I have no CDS spreads for Stifel Financial Corporation or Raymond James Financial, which are both among the most active 25 firms in terms of matched trades.

[17]For example, I lack a CDS series for First Horizon National Corporation or Jefferies Group ranking among the most active 15 of dealer firms in my sample. Moreover, neither Bloomberg nor Datastream/CMA offer CDS data on Canadian banks (e.g., Royal Bank of Canada, Scotiabank, or National Bank of Canada).

[18]When retrieving CDS data I use CBGN as Bloomberg pricing source (average prices that are computed intraday from historical snapshots taken at 5 p.m. New York time). Stale prices are the result of missing price updates (i.e., no new contributor price or no other price within 24 hours aside from the updates that initiates the intraday calculation).

(i.e., neglecting the long tail of dealers that marginally contribute to the sample).[19] Based on a dealer's rating I *impute* her CDS spread. Specifically, I compute the average CDS spread on a given day for a given rating class using the sample of dealers with both a CDS spread and a credit rating. I then map the average CDS spread per ratings class to the dealers for whom I lack a CDS spread. This seems appropriate for two reasons: first, it increases the baseline sample by more than 55,000 additional observations; and, second, it broadens the baseline sample with respect to smaller dealer firms, which often only hold a credit rating. Rating classes are defined as follows: (i) ratings Aaa to Aa3 form a class of prime and high grade ratings; (ii) ratings A1 to A3 form a class of upper medium grade ratings; (iii) ratings Baa1 to Baa3 form a class of lower medium grade ratings; (iv) ratings Ba1 to Ba3 form a class of speculative, non-investment grade ratings; (v) ratings B1 to B3 form a class of highly speculative, non-investment grade ratings; (vi) ratings Caa1 to Ca form a class of extremely speculative, non-investment grade ratings that range up to stages of imminent default with little prospect for recovery; (vii) a rating of C forms a class of defaulted dealers. Out of the 295,424 matched TRACE trades I am able to pair 231,078 (208,376) trades with a credit rating (CDS spread). Using imputed, rating-based average CDS spreads leaves me with data for 258,267 trades instead.[20]

Step 5: Computing Price and Order-Flow Differences

The dependent and independent variables in my price impact regressions are computed from price and order flow differences between trades that are strictly consecutive in time involving (potentially) *different* dealers. That is, I compare the price and order flow characteristics of a matched TRACE trade with the previous and subsequent TRACE trade.[21] In case I am pairing two matched TRACE trades I can relate the dealer-specific characteristics, CDS_t^l and MS_t^l for $l = \{i, j\}$, to the transaction price difference. Unfortunately, however, the number of consecutive matched TRACE trades is very small so I also draw on unmatched TRACE trades to compute trade-to-trade price and order flow differences.[22] Unmatched trades are anonymous with respect to dealer identities and consequently I lack dealer-specific CDS spreads and market shares. Instead, the terms $\text{CDS}_t^{\cdot}$ and $\text{MS}_t^{\cdot}$ will be

[19] Out of 207 dealers accounting for 99% of matched TRACE trades I lack ratings on 99 firms.

[20] These reflect 100% of trades of the 10 most active and 97% (98%) of trades of the 25 (50) most active dealer firms.

[21] If dealer i behind price $p^i_{t_k,n}$ is known I can compute a backward difference, $p^i_{t_k,n} - p_{t_{k-1},n}$, and a forward difference, $p_{t_{k+1},n} - p^i_{t_k,n}$, where for prices $p_{t_{k-1},n}$ and $p_{t_{k+1},n}$ respectively the dealers' identities are potentially unknown.

[22] For robustness, specification (6) in Table 2 holds the results for only matched trade pairs.

replaced with the daily (volume-weighted) sample averages $\overline{CDS}_t$ and $\overline{MS}_t$. Lastly, if the matched TRACE trade is the only trade of the day I cannot compute a within-day price difference and the observation is excluded (i.e., ca. 34% of matched trades with a dealer CDS). Excluding missing observations in the differenced order flow data, the final sample of consecutive price differences consists of 169,489 matched TRACE trades that yield 250,331 observations involving 101 dealers and 9,725 bonds of 1,922 issuers.

In addition, I also compute price and order flow differences for (potentially) non-consecutive trades that can be linked to the *same* dealer and happen within a 24-hour time window.[23] This is motivated by the balance between ensuring a sufficiently large sub-sample and avoiding too long trade time intervals that increase estimation noise. Excluding missing observations, I am left with 14,439 matched TRACE trades executed by the same dealer that yield 7,272 trade pairs, involve 3,857 different bonds of 1,144 different issuers, and are transacted by 60 different dealers.

For its size the sample with different dealer trades is used for the baseline analysis whereas the sample of same dealer trades is used for robustness. In comparison, the sample of *different* dealer pairs is considerably lager holding a much wider range of corporate bonds and therefore better reflects the average trading experience of an insurer. The sample of *same* dealer trade pairs consists of slightly larger trades in more volatile bonds that dealers tend to offset within the same day to another insurance company rather than to establish an inventory positions.

[23]Including overnight price differences yields 7,749 trade pairs instead of 6,004 trades that happen within the same trading day.

Table A-1: Steps Taken from NAIC Source File to Cleaned NAIC Sample

	July 1st, 2002 to June 30th, 2014		
	CUSIPs	**Trades**	**Trade/CUSIP**
NAIC transactions (source data)	86,909	3,182,897	37
Step (1): Eliminate based on FISD bond characteristics			
– Non-corporate bonds[1], ABS/MBS, convertibles, SLOBs, putable, and foreign currency bonds	56,316	1,218,804	
Step (2): Eliminate transactions which are not trades			
– Non-trade indicated by counterparty field entry (e.g., CALL, CONVERSION, ... etc.)	27,013	686,869	
NAIC trades (after steps 1 and 2)	26,690	1,277,224	48
Step (3): Eliminate all non-investment grade trades			
– High-yield bonds	7,442	214,874	
– Without a rating	5,503	70,892	
NAIC trades (after steps 1, 2, and 3)	19,959	991,458	50
(4) Eliminate trades with data issues			
– Missing price	422	670	
– Zero price (or zero par value)	87	108	
– Negative price	48	49	
– Negative par value	33	40	
– Price greater than 220[3]	100	148	
– Trades with par value/issuance amount *geq* 0.5	368	472	
– Trades with par value less than 100,000 dollars	11,281	111,573	
(5) Eliminated trades with timing issues			
– Trades executed on or before bond's offering date	8,800	170,750	
– Trades executed on or after bond's maturity date	1,358	1,945	
– Trades executed on weekends	1,524	2,149	
– Trades executed on SIFMA and SEC[3] holidays	1,297	1,902	
Cleaned NAIC trades (after all steps)	17,330	701,652	40

This table reports the steps taken from the NAIC source data to the Cleaned NAIC sample. All steps are taken sequentially. The ***CUSIPs*** column gives total number of unique CUSIPs eliminated from the database, the ***Trades*** column gives total number of observations eliminated. ***Trade/CUSIP*** gives the average number of trades per bond CUSIP. [1] Using bond-level characteristics such as issuer industry codes and bond type indicators (incl. Foreign government bonds and municipality bonds). [2] Price cutoff of 220 based on Asquith et al. (2013) who are computing a bond's maturity, coupons remaining, and lowest value of the treasury yield curve and taking the maximum across bonds. [3] On June 11, 2004, they day President Reagan died, the SEC declared a holiday.

Table A-2: Steps Taken from TRACE Source File to Cleaned TRACE Sample

	July 1st, 2002 to June 30th, 2014		
	CUSIPs	Trades	Trade/CUSIP
TRACE transactions (source data)	16,963	89,810,678	5,295
Step (1): Eliminate erroneous TRACE reports[1]			
– Cancellations, modifications, reversals, or duplicates	13,504	4,356,460	
Step (2): Eliminate agency and inter-dealer trades			
– Agency trades	15,962	12,859,700	
– Inter-dealer trades	16,571	39,082,374	
Step (3): Eliminate special TRACE reports			
– Reports on issuance date, non-secondary market trades, special price flags, irregular trade type, or equity-linked notes	4,492	600,842	
TRACE trades (after steps 1, 2, and 3)	16,952	32,911,302	1,941
Step (4): Eliminate all non-investment grade trades			
– High-yield bonds	2,277	3,819,219	
– Without a rating	3,400	404,953	
TRACE trades (after steps 1 to 4)	16,881	28,687,130	1,699
Step (5): Eliminate trades with data issues			
– Missing price	0	0	
– Zero price (or zero par value)	1	1	
– Negative price	0	0	
– Negative par value	0	0	
– Price greater than 220[2]	122	134	
– Trades with par value/issuance amount ≥ 0.5	645	1,254	
– Trades with par value less than 1000 dollars	16,073	19,144,068	
Step (6): Eliminated trades with timing issues			
– Trades executed on or before bond's offering date	5,449	92,282	
– Trades executed on or after bond's maturity date	84	92	
– Trades executed on weekends	163	113	
– Trades executed on SIFMA and SEC[3] holidays	5,312	19,360	
Cleaned TRACE trades (after all steps)	16,742	9,429,776	563

This table reports the steps taken from the TRACE source data to the Cleaned TRACE sample. All steps are taken sequentially. The ***CUSIPs*** column gives total number of unique CUSIPs eliminated from the database, the ***Trades*** column gives total number of observations eliminated. ***Trade/CUSIP*** gives the average number of trades per bond CUSIP. [1] I follow Dick-Nielsen (2014) and Asquith et al. (2013) with regard to erroneous reports. [2] Price cutoff of 220 based on Asquith et al. (2013) who are computing a bond's maturity, coupons remaining, and lowest value of the treasury yield curve and taking the maximum across bonds. [3] On June 11, 2004, they day President Reagan died, the SEC declared a holiday.

Table A-3: Summary Statistics (cleaned NAIC dataset)

	2002			2003			2004			2005			2006			2007			2008		
	Mean	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*
#Obs	40,138			79,914			69,962			54,087			49,393			45,075			43,243		
#Bonds	3,779			5,323			5,294			5,265			5,280			5,229			4,868		
Issuance	194	100	264	313	200	435	326	200	430	340	200	466	348	250	396	415	250	477	491	300	604
Rating	11.6	10.1	8.5	12.1	10.2	9.5	11.7	10.1	10.1	11.2	10.1	9.5	11.1	10.1	9.6	12.5	10.2	11.9	11.6	10.1	10.1
Maturity	7.0	7.0	1.3	6.7	6.8	1.3	6.3	6.5	1.6	6.1	6.3	1.6	5.9	5.9	1.5	6.0	5.9	1.3	6.1	5.9	4.0
Yield	6.9	7.0	2.1	7.1	7.0	2.1	7.0	7.0	2.2	7.1	7.0	2.2	7.0	7.0	2.1	6.7	7.0	2.3	6.5	6.0	2.3
Age	2.6	1.6	2.7	2.6	1.7	2.7	2.8	2.0	2.7	3.3	2.7	2.8	3.6	3.1	3.0	3.6	3.0	3.1	3.6	2.7	3.4
Turnover	6.9	4.8	7.6	6.6	4.7	7.1	6.5	4.8	7.6	6.1	4.5	6.6	6.0	4.4	6.8	5.6	4.1	5.8	4.9	3.5	5.3
Trd Size	2,577	1,000	4,047	2,409	1,000	3,972	2,469	1,000	4,239	2,569	1,000	4,083	2,860	1,050	4,527	2,746	1,000	4,392	2,462	1,000	4,021
#Trds p.m.	20.2	14.0	19.0	19.3	14.0	18.4	18.1	13.0	16.5	15.9	11.0	14.9	15.0	10.0	14.3	15.5	11.0	14.2	17.0	12.0	15.5
#Trds p.d.	3.2	2.0	3.7	3.0	2.0	4.0	2.6	2.0	3.1	2.0	1.0	2.4	1.9	1.0	1.6	2.1	1.0	2.5	2.4	1.0	3.0
Sells	44.5	Buy	49.7	42.8	Buy	49.5	46.0	Buy	49.8	48.7	Buy	50.0	50.3	Sell	50.0	50.2	Sell	50.0	43.2	Buy	49.5

	2009			2010			2011			2012			2013			2014			2003-2013		
	Mean	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*CAGR*
#Obs	56,551			57,297			59,634			55,936			59,503			30,919			53,973		-2.6
#Bonds	4,945			5,212			5,397			5,838			6,327			5,348			5,239		1.6
Issuance	561	329	697	603	400	747	616	450	638	545	400	560	581	400	592	720	500	709	466	306	5.8
Rating	12.2	10.1	10.4	12.2	10.2	10.4	12.2	10.2	10.8	12.6	10.2	10.7	13.0	10.2	10.8	12.3	10.2	10.7	12.0	10.1	0.6
Maturity	6.1	6.0	1.4	5.8	5.8	1.5	5.2	5.3	1.7	4.9	5.0	1.7	4.1	4.0	1.9	3.8	3.8	1.7	5.7	5.7	-4.3
Yield	6.7	7.0	2.3	6.9	7.0	2.2	7.0	7.0	2.1	7.4	7.0	1.9	7.5	8.0	1.8	7.5	8.0	1.9	7.0	7.1	0.4
Age	3.3	2.0	3.4	3.4	2.3	3.4	3.2	2.2	3.3	3.2	2.2	3.4	3.3	2.2	3.6	3.2	2.2	3.3	3.2	2.3	2.3
Turnover	4.4	3.0	5.2	3.9	2.6	4.6	3.5	2.2	4.7	3.0	1.8	4.1	2.6	1.5	3.7	2.3	1.4	3.4	4.8	3.3	-8.1
Trd Size	2,269	1,000	4,370	2,147	1,000	4,040	2,047	878	3,609	1,940	790	3,585	1,910	770	3,475	1,937	750	3,520	2,334	941	-2.1
#Trds p.m.	16.2	12.0	14.1	14.6	11.0	13.0	13.2	10.0	11.6	11.5	9.0	10.5	9.8	7.0	9.3	9.3	7.0	9.0	15.1	10.8	-6.0
#Trds p.d.	2.5	1.0	3.5	2.2	1.0	2.5	2.3	1.0	2.9	2.3	1.0	3.1	2.1	1.0	2.3	2.2	1.0	2.5	2.4	1.2	-3.1
Sells	38.7	Buy	48.7	42.5	Buy	49.4	45.8	Buy	49.8	47.8	Buy	50.0	45.1	Buy	49.8	51.6	Sell	50.0	45.9	Buy	0.5

This table reports summary statistics for the cleaned NAIC dataset (i.e., cross-sectional mean, median, and standard deviation). ***#Obs*** is the number of trades in the sample. ***#Bonds*** is the number of bonds. ***Issuance*** is a bond's issued amount (in $ millions). ***Maturity*** is the bond's time to maturity at issuance (in years). ***Yield***, reported only for fixed coupon bonds, is the bond's offering yield (in %). The last three statistics are calculated across bond issues taking each issue as one observation. The following statistics are calculated across bond issues taking each trade as one observation. ***Rating*** is a numerical translation of Moody's rating: 1=Aaa to 21=C. ***Age*** is the time since issuance (in years). ***Turnover*** is the bond's monthly trading volume as a percentage of its issued amount (in %). ***Trd Size*** is the average trade size of the bond (in $ thousands). ***#Trds p.m. (p.d.)*** is the bond's total number of trades in a month (day). ***Sells*** gives the share of customer sell orders (in %). The columns captioned with ***2003-2013*** give the average of the means and medians respectively and CAGR gives the mean annual growth rate for 2003 to 2013.

Table A-4: Summary Statistics (by most-active dealers)

	Most Active 5 Dealers			**Most Active 10 Dealers**			**Most Active 25 Dealers**			**Most Active 50 Dealers**			**All Dealers**		
	Mean	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*
#Obs	115,522			181,440			260,654			286,711			295,424		
#Bonds	9,950			10,725			11,556			11,882			12,059		
Issuance	502	300	561	513	300	573	496	300	569	484	300	561	479	300	561
Rating	12.6	10.2	10.5	12.6	10.2	10.7	12.6	10.2	10.7	12.8	10.2	10.8	12.8	10.2	10.8
Maturity	6.0	6.1	1.6	6.0	6.1	2.0	6.0	6.1	1.9	6.0	6.1	1.9	6.0	6.1	1.9
Yield	7.0	7.0	2.1	7.0	7.0	2.1	7.0	7.0	2.1	7.0	7.0	2.1	7.0	7.0	2.1
Age	3.7	2.8	3.2	3.6	2.8	3.2	3.7	2.8	3.3	3.7	2.8	3.3	3.7	2.8	3.3
Turnover	5.1	3.5	5.5	5.1	3.5	5.6	5.0	3.4	5.6	5.0	3.4	5.6	4.9	3.3	5.6
Trd Size	2,903	1,250	4,265	3,010	1,500	4,295	2,832	1,250	4,077	2,734	1,075	3,985	2,694	1,000	3,950
#Trds p.m.	14.4	10.0	13.8	14.7	10.0	14.1	14.5	10.0	13.9	14.4	10.0	13.8	14.3	10.0	13.8
#Trds p.d.	1.5	1.0	1.2	1.5	1.0	1.2	1.5	1.0	1.2	1.5	1.0	1.2	1.5	1.0	1.2
Sells	54.4	Sell	49.8	54.1	Sell	49.8	52.9	Sell	49.9	51.9	Sell	50.0	51.4	Sell	50.0

	Most Active 5 Dealers			**Most Active 10 Dealers**			**Most Active 25 Dealers**			**Most Active 50 Dealers**			**All Dealers**		
	Mean	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*	*Mean*	*Med*	*Std*
#Obs	275,121			438,955			613,256			672,293			701,652		
#Bonds	13,648			14,898			16,164			16,759			17,330		
Issuance	525	350	599	523	350	601	501	300	583	485	300	575	470	300	575
Rating	12.1	10.2	10.3	12.0	10.2	10.3	11.9	10.2	10.2	12.1	10.2	10.3	12.1	10.2	10.3
Maturity	5.6	5.7	2.3	5.6	5.8	2.1	5.6	5.8	2.1	5.6	5.8	2.0	5.6	5.8	2.0
Yield	7.1	7.0	2.1	7.1	7.0	2.1	7.0	7.0	2.1	7.0	7.0	2.1	7.0	7.0	2.1
Age	3.0	2.1	3.0	3.0	2.1	3.0	3.1	2.2	3.1	3.1	2.2	3.1	3.2	2.2	3.2
Turnover	5.1	3.4	6.0	5.1	3.4	6.1	5.0	3.3	6.1	4.9	3.2	6.1	4.9	3.2	6.0
Trd Size	2,588	1,000	4,465	2,628	1,000	4,436	2,463	1,000	4,174	2,378	1,000	4,070	2,330	1,000	4,019
#Trds p.m.	15.8	11.0	15.2	16.0	11.0	15.2	15.6	11.0	14.8	15.4	11.0	14.7	15.3	11.0	14.7
#Trds p.d.	2.6	1.0	3.6	2.6	1.0	3.4	2.4	1.0	3.1	2.4	1.0	3.0	2.4	1.0	3.0
Sells	47.4	Buy	49.9	47.2	Buy	49.9	46.7	Buy	49.9	46.0	Buy	49.8	45.6	Buy	49.8

This table reports summary statistics for the most active 5, 10, 25, and 50 dealers as well as, for comparison, for all dealers (i.e., cross-sectional mean, median, and standard deviation). The upper (lower) panel holds the data for the baseline sample of matched TRACE trades (for the cleaned NAIC dataset). ***#Obs*** is the number of trades in the sample. ***#Bonds*** is the number of bonds. ***Issuance*** is a bond's issued amount (in \$ millions). ***Maturity*** is the bond's time to maturity at issuance (in years). ***Yield***, reported only for fixed coupon bonds, is the bond's offering yield (in %). The last three statistics are calculated across bond issues taking each issue as one observation. The following statistics are calculated across bond issues taking each trade as one observation. ***Rating*** is a numerical translation of Moody's rating: 1=Aaa to 21=C. ***Age*** is the time since issuance (in years). ***Turnover*** is the bond's monthly trading volume as a percentage of its issued amount (in %). ***Trd Size*** is the average trade size of the bond (in \$ thousands). ***#Trds p.m. (p.d.)*** is the bond's total number of trades in a month (day). ***Sells*** gives the share of customer sell orders (in %).

Table A-5: Matched TRACE Trades by Year

	Cleaned NAIC and TRACE data				Sequential matching procedure			Matching Success		
Years	**NAIC CUSIPs**		**NAIC Trades**	**TRACE Trades**	**(1) Not Aggregated**	**(2) Aggregated**	**(3) Aggregated (insurer type)**	**Matched Trades**	**NAIC Trades (%)**	**TRACE Trades (%)**
2002	3,779	93.4%	40,138	335,530	12,439	2,485	173	13,461	33.5%	4.0%
2003	5,323	93.0%	79,914	705,663	27,337	4,860	214	30,430	38.1%	4.3%
2004	5,294	94.5%	69,962	598,689	26,921	4,656	183	30,365	43.4%	5.1%
2005	5,265	94.4%	54,087	537,491	24,154	3,631	84	27,166	50.2%	5.1%
2006	5,280	95.8%	49,393	497,177	23,194	3,398	75	26,183	53.0%	5.3%
2007	5,229	95.1%	45,075	474,519	19,773	2,727	103	22,102	49.0%	4.7%
2008	4,868	95.2%	43,243	598,535	17,853	2,558	156	19,857	45.9%	3.3%
2009	4,945	97.3%	56,551	944,324	22,748	3,368	199	24,701	43.7%	2.6%
2010	5,212	97.6%	57,297	893,058	25,328	3,527	145	25,347	44.2%	2.8%
2011	5,397	98.1%	59,634	938,484	26,517	3,669	185	25,001	41.9%	2.7%
2012	5,838	98.8%	55,936	1,040,891	25,018	3,350	135	22,301	39.9%	2.1%
2013	6,327	98.9%	59,503	1,110,482	27,173	3,529	126	20,185	33.9%	1.8%
2014	5,348	99.3%	30,919	544,362	14,120	1,681	55	8,325	26.9%	1.5%
Total			**701,652**	**9,219,205**	**292,575**	**43,439**	**1,833**	**295,424**	**42.1%**	**3.2%**

This table holds the number of matches within the cleaned NAIC and the cleaned TRACE data on a yearly basis. The ***NAIC CUSIPs*** column gives the number of unique CUSIPs found in the cleaned NAIC data where the coverage in TRACE (in %) for the particular year is reported in brackets. Columns ***NAIC Trades*** and ***TRACE Trades*** give the trades within each cleaned dataset for the *set of NAIC CUSIPs*. The column ***(1) Non Aggregated*** reports matches of non-aggregated (i.e., as reported) NAIC trades with TRACE trades; **(2) Aggregated** reports matches of aggregated NAIC trades per bond CUSIP when trade execution date, counterparty, and buy/sell indicator are identical and the difference in price is smaller or equal to one cent (i.e., $\leq |0.01|$ on a \$100-par-basis); **(3) Aggregated (insurer type)** reports matches of aggregated NAIC trades where grouping includes the variable *insurer type*. The matching steps are taken sequentially. In case multiple NAIC trades are matched on the same TRACE trade these trades are deleted if the counterparty variable differs. Otherwise I give preferences to the trade of the earliest matching step (i.e., (1) before (2) before (3)). In case one NAIC trade matches on multiple TRACE trades the trade is discarded (i.e., not considered a match). The columns labelled ***Matching Success*** report the number of matched trades and the ratio of NAIC and TRACE trades that have been matched respectively.

A.2 Model Appendix

This section offers a step-by-step derivation of the price impact regression model given in Section 2 of the paper "*Middlemen Matter: Corporate Bond Market Liquidity and Dealer Inventory Funding*".[24] Let t_k index the date and time of a trade in bond n where t for $t = 1, ..., T$ refers to the particular trading day and k for $k = 1, ..., K$ refers to the execution time.[25] The reduced-form transaction price model, for dealer i's observed transaction price, $p^i_{t_k,n}$, holds three key ingredients: First, the unobservable fundamental value of the bond, $p^*_{t_k,n}$, in absence of transaction costs. Second, the quote midpoint, $m^i_{t_k,n}(p^*_{t_{k-1},n}, I^i_{t_k,n})$, representing dealer i's valuation of the fundamental value process, $p^*_{t_{k-1},n}$, and factoring in her inventory costs for inventory level, $I^i_{t_k,n}$, just before executing the order at time t_k. Third, $\frac{1}{2}S^i_{t_k,n}\left(\lambda_{t_k}, \gamma^i_{t_k}, \beta^i_{t,n}\right)$, reflecting half of the bid-ask spread at time t_k and being a function of three liquidity frictions: order-processing costs ($\gamma^i_{t_k}$), adverse selection costs (λ_{t_k}), and inventory costs ($\beta^i_{t,n}$). The transaction price is then given by:

$$p^i_{t_k,n} = m^i_{t_k,n}\left(p^*_{t_{k-1},n}, I^i_{t_k,n}\right) + \frac{1}{2}S^i_{t_k,n}\left(\lambda_{t_k}, \gamma^i_{t_k}, \beta^i_{t,n}\right) d_{t_k,n} \tag{A.2.1}$$

where $d_{t_k,n}$ indicates whether a trade is a customer buy order at the ask (i.e., $d_{t_k,n} = 1$) or a customer sell order at the bid (i.e., $d_{t_k,n} = -1$).

A three-way decomposition of the bid-ask spread rests on the implicit assumption that the fundamental value of the security is affected by surprises in the order flow, which requires an assumption about the underlying process of trade flows. I exploit the fact that market orders, for various reasons, can indeed be serially correlated. In inventory models quote changes affect the subsequent arrival rate of incoming orders (Stoll, 1978; Ho and Stoll, 1981). Such behavior induces negative serial correlation in market orders and quote changes (e.g., documented by Friewald and Nagler (2015)). To reflect serial correlation in market orders, order flow is assumed to follow a first-order autoregressive process given by equation (A.2.2)

$$q_{t_k,n} = \phi q_{t_{k-1},n} + \eta_{t_k,n} \tag{A.2.2}$$

[24] The model builds on Section 5 (pp. 1011) "*Three-Way Decomposition of the Spread Based on Induced Serial Correlation in Trade Flows*" in Huang and Stoll (1997). My notation differs from that of Huang and Stoll (1997) though.

[25] The set of equations represent three separate and sequential events subsumed under the time subscript t_k. In comparison to Huang and Stoll (1997) I use a different timing convention: $p^*_{t_k,n}$ includes the information revealed by the transaction at time t_k (specifically the private information revealed by the order flow); $m^i_{t_k,n}(\cdot)$ incorporates the dealer's aggregate inventory, $I^i_{t_k,n}$, just before the transaction at time t_k.

where ϕ is expected to be negative, $\eta_{t_k,n}$ is assumed to be white noise (i.e., $\mathrm{E}[\eta_{t_k,n}] = 0\ \forall t_k$ and $\mathrm{E}[\eta_{t_k,n}, \eta_{t_s,n}] = 0\ \forall t_k \neq t_s$), and $q_{t_k,n} = d_{t_k,n}|q_{t_k,n}|$ represents the signed trade size at time t_k.

The conditional expectation of trade size at time t_{k-1} is given by $\mathrm{E}[q_{t_k,n}|\Omega_{t_{k-1}}] = \phi q_{t_{k-1},n}$ where $\Omega_{t_{k-1}}$ reflects all relevant information available before the t_k^{th} transaction. Re-arranging equation (A.2.2) yields the following expression for the order flow innovation

$$\eta_{t_k,n} \equiv q_{t_k,n} - \mathrm{E}[q_{t_k,n}|\Omega_{t_{k-1}}] = \phi q_{t_{k-1},n} \ . \tag{A.2.3}$$

Based on the assumption that order flow is correlated and the market knows equation (A.2.2), the unobservable fundamental value, $p^*_{t_k}$, follows equation (A.2.4)

$$p^*_{t_k,n} = p^*_{t_{k-1},n} + \lambda_1 \left(q_{t_k,n} - \phi q_{t_{k-1},n}\right) + \epsilon_{t_k,n} \tag{A.2.4}$$

where λ_1 represents the percentage of the half-spread attributable to adverse selection costs, and $\epsilon_{t_k,n}$ is a serially uncorrelated public information shock (i.e., $\mathrm{E}[\epsilon_{t_k,n}] = 0$ and $\mathrm{E}[\epsilon_{t_k,n}, \epsilon_{t_s,n}] = 0 \forall t_k$, $\forall t_k \neq t_s$). Equation (A.2.4) implies that the price process resembles a random walk dependent on the trading process through the unexpected component in order flow. This entails a tacit assumption about the rationality of dealer i as any trending serial correlation could be exploited.

Subtracting $p^*_{t_{k-1},n}$ from both sides of equation (A.2.4) yields the change in the "true" price process, $p^*_{t_k,n} - p^*_{t_{k-1},n}$. Defining the latter as $\Delta p^*_{t_k,n}$ where Δ represents the first difference operator (for trades within the same trading day t) this yields

$$\Delta p^*_{t_k,n} = \lambda_1 \left(q_{t_k,n} - \phi q_{t_{k-1},n}\right) + \epsilon_{t_k,n} \ . \tag{A.2.5}$$

Given the definition of (A.2.4), changes in $p^*_{t_k,n}$ are induced either by the price impact of private information revealed in order flow innovations (the first term), or due to the arrival of unexpected public information releases (the last term). Prices changes in the fundamental value as given by equation (A.2.5) remain serially uncorrelated and unpredictable. Consider, for example, the expectation of $\Delta p^*_{t_k,n}$ conditional on $\Omega_{t_{k-1}}$

$$\mathrm{E}[\Delta p^*_{t_k,n}|\Omega_{t_{k-1}}] = \mathrm{E}[\lambda_1 \left(q_{t_k,n} - \phi q_{t_{k-1},n}\right) + \epsilon_{t_k,n}|\Omega_{t_{k-1}}] = \mathrm{E}[\lambda_1 \eta_{t_k,n} + \epsilon_{t_k,n}|\Omega_{t_{k-1}}] = 0 \ .$$

The unobservable hypothetical construct of the fundamental value determines dealer i's valuation of the bond. At the same time, she also factors in her inventory holding costs to induce inventory-equilibrating trades. Consequently, the midpoint, just before executing the order and observing the order flow innovation at time t_k, is related to the fundamental value according to the following equation

$$m^i_{t_k,n} = p^*_{t_{k-1},n} + \epsilon_{t_k,n} - \beta^i_{t,n} I^i_{t_k,n} \tag{A.2.6}$$

where $\beta^i_{t,n}$ reflects inventory costs on trading day t (assumed to be constant over the trading day), and $I^i_{t_k,n}$ resembles dealer i's aggregate inventory in the bond just before the transaction at time t_k. As the spread is assumed to be symmetric the midquote equation is valid for both trades at the ask and the bid.

Equation (A.2.6) captures that only dealer i's actual inventory exposure and not the unexpected portion matters. That is, inventory costs arise only when inventory is acquired (even if the inventory change was expected) and do not play a role if inventory is not acquired (even if the lack of inventory change was unexpected). Midquote adjustments for inventory reasons thereby depend on actual trades, not trade surprises. This distinction allows the separate identification and estimation of the inventory and the adverse selection component.

I differentiate between several inventory subcomponents where the linear specification of $\beta^i_{t,n}$ takes the following form:

$$\beta^i_{t,n} = \beta_0 + \sum_{r=2}^{10} \beta_{0,r}\, \mathrm{CR}^r_{t,n} + \beta_1\, \mathrm{SYS\_RV}_{t,n} + \beta_2\, \mathrm{IDIO\_RV}_{t,n} + \beta_3\, \mathrm{TED}_t + \beta_4\, \mathrm{CDS}^i_t \tag{A.2.7}$$

where $\mathrm{CR}^r_{t,n}$ is a dummy variable referring to a bond's credit rating and equal to 1 in case bond n holds rating r for $r = 2, ..., 10$ (where Moody's investment-grade ratings range from 1=Aaa to 10=Baa3), TED_t refers to the TED spread on day t (i.e., the difference between the three-month LIBOR and the three-month T-bill interest rates), $\mathrm{SYS\_RV}_{t,n}$ ($\mathrm{IDIO\_RV}_{t,n}$) is the bond's realized systematic (idiosyncratic) volatility using a 90-day rolling window, and CDS^i_t is dealer i's CDS spread on trading day t.

A bond's 90-day total volatility is based on all cleaned customer-dealer TRACE trades and defined as $\mathrm{TOTAL\_RV}_{t,n} = \sqrt{\frac{1}{\sum_{j=1}^{90} d_{t-j,n}} \sum_{j=1}^{90} d_{t-j,n}\, r^{e,2}_{t-j,n}}$ where $r^{e,2}_{t-j,n} = \ln(p_{t-j,n}/p_{t-j-1,n}) - r^f_{t-j}$ is a log-return in excess of the risk-free rate, r^f_{t-j} (over the same holding period), and $d_{t-j,n}$ is a dummy that is zero in case of non-trading days (e.g., weekends or trading holidays) or stale prices. In case two consecutive price observations are not available (e.g., on the first day of trading after a period of no trading) I compute a return using the last available (stale) price. In case of multiple trades per day I first compute the weighted price of the day using trade size. Following the standard practice in the literature the mean correction of returns is omitted. To ensure that the dependent variable is not affected by potential contemporaneous correlation with the realized volatility I only include lagged returns.

Using a one-factor model, I break down $\text{TOTAL\_RV}_{t,n}$ into systematic volatility ($\text{SYS\_RV}_{t,n}$) and idiosyncratic volatility ($\text{IDIO\_RV}_{t,n}$). That is, I regress bond returns on returns from an equally-weighted market portfolio of bond's that hold a bond credit rating of 7 (i.e., A3) and a maturity of 10 years. The bonds applicable for the market portfolio come from the list of all CUSIPs within the cleaned NAIC sample. This regression equation follows:

$$r^e_{t,n} = \alpha_n + \beta_n r^e_{t,\text{MKT}} + \epsilon_{t,n}$$

where $r^e_{t,n}$ is a bond's excess return; $r^e_{t,\text{MKT}}$ is the market portfolio's excess return (over the same holding period); α_n is the intercept; and $\epsilon_{t,n}$ is the regression residual. Based on the residuals and using the most recent 90 days of data for each bond, I define the standard deviation of the model residuals as the realized idiosyncratic volatility, $\text{IDIO\_RV}_{t,n} = \sqrt{\frac{1}{\sum_{j=1}^{90} d_{t-j,n}} \sum_{j=1}^{90} d_{t-j,n}\, \epsilon^2_{t-j,n}}$. The realized systematic volatility is then given by the following formula $\text{SYS\_RV}_{t,n} = \sqrt{\text{TOTAL\_RV}^2_{t,n} - \text{IDIO\_RV}^2_{t,n}}$.

Taking the first-difference of the midquote, $m^i_{t_k,n} - m^i_{t_{k-1},n}$, from equation (A.2.6) one obtains

$$\Delta m^i_{t_k,n} = \Delta p^*_{t_{k-1},n} + \Delta\epsilon_{t_k,n} - \beta^i_{t,n}\Delta I^i_{t_k,n} \tag{A.2.8}$$

Substituting the change in the fundamental price from equation (A.2.5) allows me to drop the lagged public news shock $\epsilon_{t_{k-1},n}$ so that the equation simplifies to

$$\Delta m^i_{t_k,n} = \lambda_1 \left(q_{t_{k-1},n} - \phi q_{t_{k-2},n}\right) + \beta^i_{t,n} q_{t_{k-1},n} + \epsilon_{t_k,n} \tag{A.2.9}$$

where I rewrite the evolution in dealer i's inventory as $\Delta I^i_{t_k,n} = -q_{t_{k-1},n}$ (i.e., by definition or market clearing the change in inventories mirrors the order flow). Equation (A.2.9) implies that quote changes reflect the information revealed by the surprise in the last order as well as the inventory costs of the last trade.

Unlike the expected change in the fundamental value, the expected change in the quote midpoint can be predicted on the basis of past trades. Consider the expectation of $\Delta m^i_{t_k,n}$ conditional on $\Omega_{t_{k-2}}$. That is, when $m^i_{t_{k-1},n}$ has been formed (and thus $q_{t_{k-2},n}$ been observed) but before $q_{t_{k-1},n}$ has been realized:

$$\begin{aligned} \mathrm{E}[\Delta m^i_{t_k,n}|\Omega_{t_{k-2}}] &= \mathrm{E}[\lambda_1 \left(q_{t_{k-1},n} - \phi q_{t_{k-2},n}\right) + \beta^i_{t,n} q_{t_{k-1},n} + \epsilon_{t_k,n}|\Omega_{t_{k-1}}] \\ &= \mathrm{E}[\lambda_1 \left(\phi q_{t_{k-2},n} - \phi q_{t_{k-2},n}\right) + \beta^i_{t,n} q_{t_{k-1},n} + \epsilon_{t_k,n}|\Omega_{t_{k-1}}] = \beta^i_{t,n} \phi q_{t_{k-2},n} \end{aligned} \tag{A.2.10}$$

This means that the *expected* midpoint change does not depend on the adverse selection component while the *observed* midpoint change does.

The transaction price at time t_k is the midquote plus or minus half the bid-ask spread depending on whether the incoming order is a buy or a sell and given by the following equation

$$p^i_{t_k,n} = m^i_{t_k,n} + \gamma^i_{t_k,n} d_{t_k,n} + \lambda_1(q_{t_k,n} - \phi q_{t_{k-1},n}) + \beta^i_{t,n} q_{t_k,n} \tag{A.2.11}$$

To distinguish potential subcomponents in the order-processing cost component, $\gamma^i_{t_k}$, I impose a linear specification that takes trade size and a dealer's market share into account. It takes the following form

$$\gamma^i_{t_k} = \gamma_0 + \gamma_1 \,\mathrm{MS}^i_t + \gamma_2 \,|q_{t_k,n}| \tag{A.2.12}$$

where γ_0 captures round-trip costs per \$100-par, γ_1 captures potential markups or discounts due to a dealer's market share (i.e., MS^i_t is defined as the ratio of trades per dealer per month to the total number of trades per month), and γ_2 captures potential markups or discounts associated with (absolute) trade size (e.g., potential "quantity discounts" which have spreads decreasing with order size). A dealer's market share stands as a proxy for dealer size and overall presence in the market.

Taking the first-difference, $p^i_{t_k,n} - p^i_{t_{k-1}}$, of equation (A.2.11) yields

$$\Delta p^i_{t_k,n} = \Delta m^i_{t_k,n} + \gamma^i_{t_k} \Delta d_{t_k,n} + \lambda_1 \left(\Delta q_{t_k,n} - \phi \Delta q_{t_{k-1},n}\right) + \beta^i_{t,n} \Delta q_{t_k,n} \tag{A.2.13}$$

Combining it with equation (A.2.9) the terms $\lambda_1(q_{t_{k-1},n} - \phi q_{t_{k-2}})$ and $\beta^i_{t,n} q_{t_{k-1},n}$ cancel out and the the equation simplifies to the following specification:

$$\begin{aligned}
p^i_{t_k,n} - p^i_{t_{k-1},n} = {} & \gamma_0 \Delta d_{t_k,n} + \gamma_1 \mathrm{MS}^i_t \Delta d_{t_k,n} + \gamma_2 \Delta q_{t_k,n} \\
& + \lambda_1 \left(q_{t_k,n} - \phi q_{t_{k-1},n} \right) \\
& + \beta_0 q_{t_k,n} + \sum_{r=2}^{10} \beta_{0,r} \, \mathrm{CR}^r_{t,n} q_{t_k,n} \\
& + \beta_1 \mathrm{SYS_RV}_{t,n} q_{t_k,n} + \beta_2 \mathrm{IDIO_RV}_{t,n} q_{t_k,n} \\
& + \beta_3 \mathrm{TED}_t q_{t_k,n} + \beta_4 \mathrm{CDS}^i_t q_{t_k,n} + \epsilon_{t_k,n}
\end{aligned} \tag{A.2.14}$$

Now, accounting for dealer identities we treat realized transaction prices as the outcome of two different dealers. Then, equations (A.2.6) and (A.2.11) need to be adapted while the unobservable fundamental value, $p^*_{t_k,n}$ does not change. To exemplify this, consider a bond trading twice on the same trading day t at times t_k and time t_{k-1}. Furthermore, let transaction price $p^j_{t_{k-1},n}$ charged by dealer j be followed by transaction price $p^i_{t_k,n}$ charged by dealer i. Then, for instance, the midquote equation and the transaction price for dealer j are given by

$$m^j_{t_{k-1},n} = p^*_{t_{k-2},n} + \epsilon_{t_{k-1},n} - \beta^j_{t,n} I^j_{t_{k-1},n} \tag{A.2.15}$$

$$p^j_{t_{k-1},n} = m^j_{t_{k-1},n} + \gamma^j_{t_{k-1}} d_{t_{k-1},n} + \lambda_1 \left(q_{t_{k-1},n} - \phi q_{t_{k-2},n} \right) + \beta^j_{t,n} q_{t_{k-1},n} \tag{A.2.16}$$

where $I^j_{t_{k-1},n}$ is the aggregate inventory before the transaction at time t_{k-1}. Similarly, for dealer i the midquote and transaction price equation are given by

$$m^i_{t_k,n} = p^*_{t_{k-1},n} + \epsilon_{t_k,n} - \beta^i_{t,n} I^i_{t_k,n} \tag{A.2.17}$$

$$p^i_{t_k,n} = m^i_{t_k} + \gamma^i_{t_k} d_{t_k,n} + \lambda_1 \left(q_{t_k,n} - \phi q_{t_{k-1},n} \right) + \beta^i_{t,n} q_{t_k,n} \tag{A.2.18}$$

where $I^i_{t_k,n}$ are their aggregate inventories before the transaction at time t_k.

The transaction price difference involves the change in dealer-specific midquotes. Subtracting equations (A.2.17) and (A.2.15) yields

$$\begin{aligned} m^i_{t_k,n} - m^j_{t_{k-1},n} = {} & \lambda_1 \left(q_{t_{k-1},n} - \phi q_{t_{k-2},n}\right) + \epsilon_{t_k,n} \\ & - \beta_0 \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) - \beta_{0,r}\, \mathrm{CR}^r_{t,n} \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) \\ & - \beta_1 \mathrm{SYS\_RV}_{t,n} \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) - \beta_1 \mathrm{IDIO\_RV}_{t,n} \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) \\ & - \beta_3 \mathrm{TED}_t \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) - \beta_4 \left(\mathrm{CDS}^i_t I^i_{t_k,n} - \mathrm{CDS}^j_t I^j_{t_{k-1},n}\right) \end{aligned} \tag{A.2.19}$$

The difference between transaction prices, $p^i_{t_k,n} - p^j_{t_{k-1},n}$, given in equations (A.2.18) and (A.2.16) is then given by

$$\begin{aligned} p^i_{t_k,n} - p^j_{t_{k-1},n} = {} & m^i_{t_k,n} - m^j_{t_{k-1},n} + \gamma_0 \Delta d_{t_k,n} + \gamma_1 \left(\mathrm{MS}^i_t d_{t_k,n} - \mathrm{MS}^j_t d_{t_{k-1},n}\right) + \gamma_2\, \Delta q_{t_k,n} \\ & + \lambda_1 \left(\Delta q_{t_k,n} - \phi \Delta q_{t_{k-1},n}\right) \\ & + \beta_0 \Delta q_{t_k,n} + \beta_{0,r}\, \mathrm{CR}^r_{t,n} \Delta q_{t_k,n} \\ & + \beta_1 \mathrm{SYS\_RV}_{t,n} \Delta q_{t_k,n} + \beta_1 \mathrm{IDIO\_RV}_{t,n} \Delta q_{t_k,n} \\ & + \beta_3 \mathrm{TED}_t \Delta q_{t_k,n} + \beta_4 \left(\mathrm{CDS}^i_t q_{t_k,n} - \mathrm{CDS}^j_t q_{t_{k-1},n}\right) \end{aligned} \tag{A.2.20}$$

When combining equations (A.2.19) and (A.2.20) we make use of the dealers' change in inventories $-q_{t_k,n} = I^i_{t_{k+1},n} - I^i_{t_k,n}$ and $-q_{t_{k-1},n} = I^j_{t_k,n} - I^j_{t_{k-1},n}$ to factor out inventory subcomponents. Then, the term $\lambda_1 (q_{t_{k-1},n} - \phi q_{t_{k-2},n})$ can be dropped and the expression simplifies to

$$\begin{aligned} p^i_{t_k,n} - p^j_{t_{k-1},n} = {} & \gamma_0 \Delta d_{t_k,n} + \gamma_1 \left(\mathrm{MS}^i_t d_{t_k,n} - \mathrm{MS}^j_t d_{t_{k-1},n}\right) + \gamma_2 \Delta q_{t_k,n} \\ & + \lambda_1 \left(q_{t_k,n} - \phi q_{t_{k-1},n}\right) \\ & + \beta_0 \Delta q_{t_k,n} - \beta_0 \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) \\ & + \sum_{r=2}^{10} \beta_{0,r}\, \mathrm{CR}^r_{t,n} \Delta q_{t_k,n} - \sum_{r=2}^{10} \beta_{0,r}\, \mathrm{CR}^r_{t,n} \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) \\ & + \beta_1 \mathrm{SYS\_RV}_{t,n} \Delta q_{t_k,n} - \beta_1 \mathrm{SYS\_RV}_{t,n} \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) \\ & + \beta_2 \mathrm{IDIO\_RV}_{t,n} \Delta q_{t_k,n} - \beta_2 \mathrm{IDIO\_RV}_{t,n} \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) \\ & + \beta_3 \mathrm{TED}_t \Delta q_{t_k,n} - \beta_3 \mathrm{TED}_t \left(I^i_{t_k,n} - I^j_{t_{k-1},n}\right) \\ & + \beta_4 \left(\mathrm{CDS}^i_t q_{t_k,n} - \mathrm{CDS}^j_t q_{t_{k-1},n}\right) - \beta_4 \left(\mathrm{CDS}^i_t I^i_{t_k,n} - \mathrm{CDS}^j_t I^j_{t_{k-1},n}\right) + \epsilon_{t_k,n} \end{aligned} \tag{A.2.21}$$