

Corporate Bond Price Reversals

Alexey Ivashchenko*

November 1, 2019

JOB MARKET PAPER.

THE MOST RECENT VERSION IS [HERE](#).

Abstract

I demonstrate that investors trade U.S. corporate bonds not only for liquidity reasons but also on private information. Bond dealers let less-informed investors provide liquidity to informed traders and are not adversely selected. I obtain these results by contrasting corporate bond price reversals in bonds with different information asymmetry, trading volume, and dealers' capital commitment. I find strong price reversals that become less pronounced following high-trading-volume days. The effect is the strongest for bonds with high information asymmetry, and when dealers' end-of-day inventory does not change. The results suggest that information reveals itself in prices on high-volume days when dealers do not accept overnight inventory risk. The findings are in line with the predictions of a theoretical model in which investors trade both for liquidity reasons and on private news that arrive independently of changes in inventory. I further show that realized bid-ask spreads are not wide enough to negate reversal profits of high-asymmetry bonds. Such reversal portfolios earn 3% per year after trading cost adjustment. By connecting low market transparency with high non-fundamental price volatility, the paper also contributes to the ongoing policy debate.

JEL classification: G12, G14.

Keywords: corporate bonds, trading volume, reversal, informed trading, dealer inventory

*University of Lausanne and Swiss Finance Institute. I thank Michael Rockinger (advisor), Vincent Bogausslavsky, Pierre Collin-Dufresne, Robert Czech, Thomas Geelen, Jakub Hajda, Erik Hapnes, Robert Kosowski, Roni Michaely, Artem Neklyudov, Maureen O'Hara, Norman Schürhoff, Karamfil Todorov, and Johan Walden for their helpful suggestions. I also thank the participants of the SFI 2019 Job Market Workshop, HEC Paris 2019 PhD Workshop, Swiss Finance Institute Research Days in Gerzensee, brownbag seminars at the Bank of England, the University of Lausanne, the University of Geneva, and the University of St. Gallen, for insightful comments and feedback. All remaining errors are my own.

Mail: Alexey.Ivashchenko@unil.ch; Web: ivasche.com

I. Introduction

Sophisticated investors used to own a substantial fraction of U.S. corporate bonds around the global financial crisis of 2008–2009. Figure 1 shows that hedge funds’ corporate bond holdings stood at around 40% of the combined holdings of insurance companies, pension funds, mutual funds, and ETFs around the time of the crisis. Ten years later, this ratio is four times lower. Citi, one of the biggest corporate bond dealers, states that ‘market diversity has fallen significantly, the buyer base has become more homogeneous’ ([Citi 2018](#)). As ‘smart money’ was leaving the market, both industry participants and academics expressed concerns that the price discovery mechanism in corporate bonds might be impaired. The market has been serving primarily large institutions trading for liquidity reasons; information-driven trading has become scarce.¹

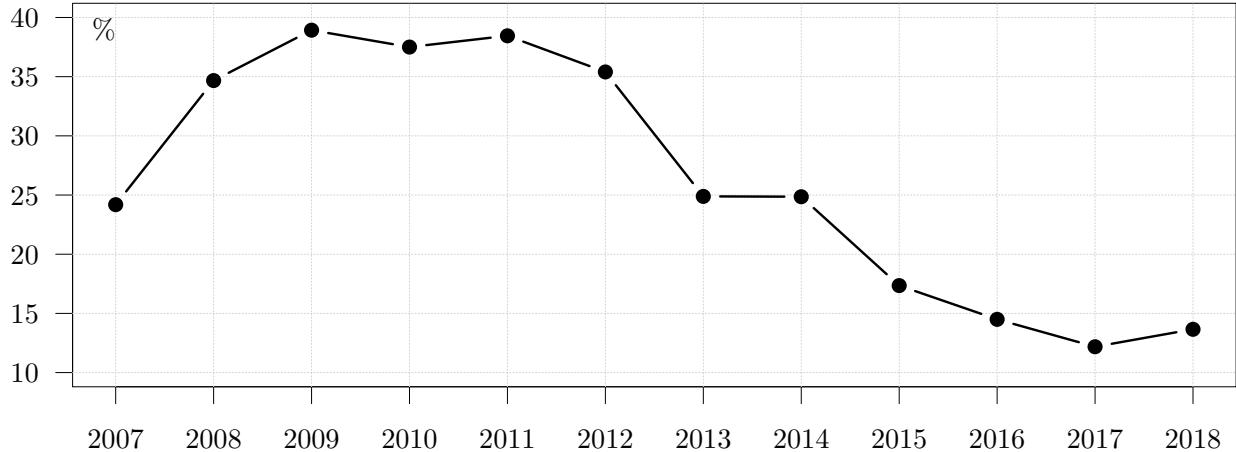


Figure 1. Hedge funds’ corporate bond holdings, in % of the combined holdings of insurance companies, pension funds, mutual funds, and ETFs. I use U.S. Flow of Funds (FF) data to calculate the ratio. The FF data do not separate hedge funds, but the industry tradition is to interpret households’ corporate bond holdings as the ones dominated by hedge funds.

¹Business cycle, tighter regulation of dealer banks, and the emergence of alternative credit trading venues all contributed to the flight of ‘smart money’ away from corporate bonds. As BlackRock writes, ‘some investors have migrated risk exposure from the cash bond market to standardized derivatives to the extent they have the flexibility to do so from a legal, regulatory, operational, and investment policy perspective’ ([BlackRock 2018](#)). Simultaneously, some scholars argue that even bond short-sellers are not trading corporate bonds on information (see, for instance, [Asquith, Au, Covert, and Pathak 2013](#)). [Berndt and Zhu \(2018\)](#) provide a model that links higher dealer inventory costs with lower market efficiency post-crisis.

In this paper, I demonstrate that, despite these concerns, there is strong empirical evidence that investors still trade corporate bonds not only for liquidity reasons but also on information. Information-driven trading is more likely in bonds with fewer mutual fund owners, fewer dealers, no actively traded CDS contracts, lower outstanding amounts, and when bond issuers are smaller firms with more volatile stocks. I call such bonds high-information-asymmetry bonds. The paper claims that bond dealers are aware of information-based trading and manage to avoid informed flows. When approached by a client who wants to trade, dealers choose whether to provide liquidity themselves or to find another investor who wants to trade in the opposite direction and let him or her provide liquidity.² I demonstrate that the latter rather than the former happens for high-information-asymmetry bonds.

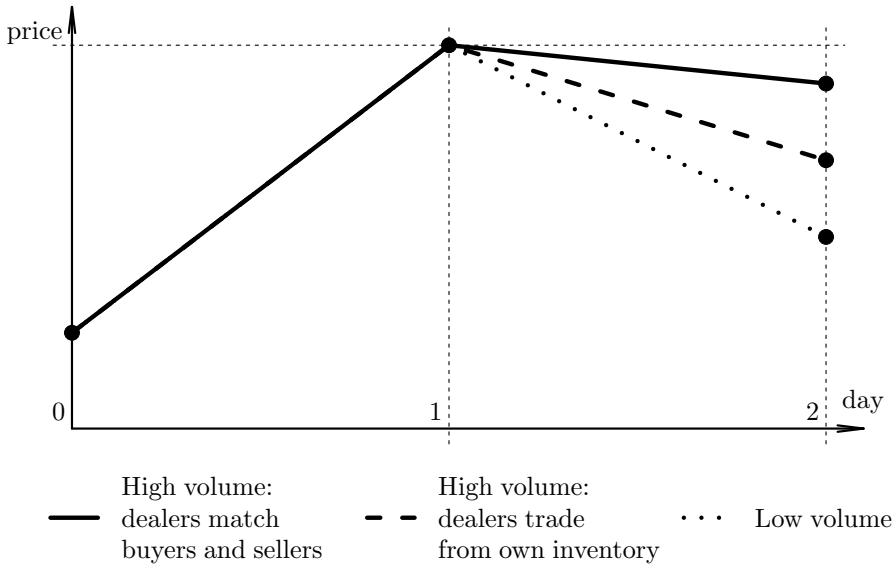


Figure 2. Stylized price reversal paths for a high-information-asymmetry bond. On day 1, the trading volume is either low or high. The solid line shows a reversal path on a high-volume day when dealers' end-of-day inventory (in this particular bond) does not change, and dealers buy from some investors as much as they sell to other investors. The dashed line refers to when trading volume on day 1 is high, and dealers trade a lot from their inventory. The 'Low volume' dotted line represents the average reversal path. For comparison purposes, I assume that the price change on day 1 is the same in all three cases.

²The dealer nevertheless executes both trades, but such pre-arranged transactions close fast, and bonds do not stay on dealer's books for longer than several minutes.

I obtain these results by contrasting corporate bond price reversals (measured as the first autocorrelation of returns) following days with different trading volumes and dealers' capital commitment. What is the link between price reversals and trading motives? Liquidity trading (non-informational trading) generates reversals, which represent remuneration for liquidity providers. Reversals tend to be less pronounced following high-trading-volume days. On such days, price changes are more persistent because trading is partly driven by private information.³ Price changes are the most persistent following high-volume days when dealers buy from some investors as much as they sell to other investors (and dealers' end-of-day inventory does not change). Figure 2 shows the stylized reversal paths I obtain for a typical high-information-asymmetry bond. Reversals are, on average, strong, but price changes become more persistent as trading volume increases, especially if dealers only match buyers and sellers and do not accept overnight inventory risk. The more persistent price changes are, the more likely it is that trading is information-motivated.

Formally, my empirical analysis proceeds in two steps. In the first step, I use TRACE data from years 2010–2017 aggregated to the daily frequency to estimate the following volume-return relationship for individual corporate bonds:⁴

$$R_{t+1} = \beta_0 + \underbrace{(\beta_1 + \beta_2 \cdot \text{Inventory-neutral volume}_t + \beta_3 \cdot |\Delta\text{Inventory}|_t)}_{\text{Return autocorrelation}} R_t + \epsilon_{t+1}, \quad (1)$$

Above, R_{t+1} stands for total corporate bond return on day $t+1$. Inventory-neutral volume is the volume of investors' purchases from dealers matched by investors' sales to dealers within business day t ; it does not add to dealers' aggregate end-of-day inventory in this bond. The difference between investors' purchases and sales is the change in dealers' inventory on day t : it stays on dealers' books until day $t+1$. High trading volume on day t can be due to high inventory-neutral volume, or a big change in dealers' inventory, or both.⁵ In (1), β_1 measures the reversal on a low-volume day, while β_2 and β_3 capture how the reversal changes

³I assume that new public information affects prices without inducing abnormally high trading volumes.

⁴I require the bonds to be traded frequently enough to be included in the sample.

⁵In this paper, I do not take into account inter-dealer trading volumes. If there are only inter-dealer trades on day t , both trading volume measures in equation (1) are zero.

following high-volume days with different dealers' capital commitment. The volume-return relationship (1) stems from a theoretical model where risk-averse investors trade corporate bonds with each other for either liquidity or informational reasons, and inventory fluctuates independently of news arrival.

In the second step, I run a cross-sectional regression of estimated volume-return coefficients $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ on information asymmetry proxies controlling for bond illiquidity, riskiness, and volume persistence. My information asymmetry proxies are the number of mutual funds that hold the bond, the number of dealers who intermediate trades in the bond, the size of the issue and the issuer, the availability of an actively traded CDS contract on the bond issuer, and issuer's stock return volatility. Larger values for all proxies except for stock volatility are associated with lower information asymmetry.

I find that $\hat{\beta}_1$ is negative. Bond prices tend to revert following low-volume days. For a typical high-asymmetry bond, $\hat{\beta}_1$ stands at around -0.4; if the price increases by 100 b.p. on a low-volume day, it falls by 40 b.p. the next day. For the same high-asymmetry bond, $\hat{\beta}_2$ is positive. For every additional standard deviation of inventory-neutral trading volume, return autocorrelation increases (reversal reduces) by 0.1. $\hat{\beta}_3$ is about two times smaller than $\hat{\beta}_2$ for the high-asymmetry bond. The results suggest that bond price changes are the most persistent when trading volumes are high, but dealers are reluctant to trade from their inventory capacity. Furthermore, I find that $\hat{\beta}_1$ decreases, $\hat{\beta}_2$ increases, and $\hat{\beta}_3$ does not change as information asymmetry grows in the cross-section of bonds.⁶ These findings suggest that information-motivated trading in corporate bonds does exist, and it most likely occurs on high-volume days when dealers are only matching buyers and sellers and do not accept additional inventory risk.

This paper further argues that the long part of the bond reversal investment strategy, constructed on higher-asymmetry bonds, delivers higher risk-adjusted returns after trading cost adjustment. Between October 2005 and June 2017, the long-only monthly re-balanced

⁶These results hold for both investment-grade and high-yield bonds, and within bonds of the same issuer (for the issuers with many bonds outstanding).

reversal portfolio on high-information-asymmetry bonds earned 2.8% annualized return after trading cost adjustment, which is 1.5 p.p. above the corporate bond market and the long-reversal return on low-asymmetry bonds. These results suggest that, even when illiquidity is taken into account, reversal returns are high. An investor implementing a bond reversal strategy in practice may further refine it using information asymmetry proxies to obtain even better performance.

My paper contributes to several streams of corporate bond literature. The paper discusses the impact of private information on corporate bond price reversals and, with this regard, extends a traditional explanation of reversals based on illiquidity stemming from OTC market frictions. [Duffie, Gârleanu, and Pedersen \(2005\)](#) present a theoretical framework where OTC market frictions drive illiquidity; [Friewald and Nagler \(2019\)](#) provide supporting empirical evidence from the corporate bond market. I demonstrate that in the cross-section of bonds with similar illiquidity, the reversals further depend on information asymmetry. In related work, [Bao, Pan, and Wang \(2011\)](#) study the cross-sectional determinants of negative bond return covariance in pre-crisis years. They find that return covariance is above and beyond the levels that can be explained by bid-ask spreads but do not link the unexplained part directly to information asymmetry.⁷

[Chordia, Goyal, Nozawa, Subrahmanyam, and Tong \(2017\)](#), [Bali, Subrahmanyam, and Wen \(2018\)](#), and [Bai, Bali, and Wen \(2019\)](#) also discuss an empirical link between corporate bond price reversals and illiquidity in the context of pricing the cross-section of corporate bonds. The papers find that one-month lagged return is the strong return predictor in the cross-section of corporate bonds. [Chordia et al. \(2017\)](#) show, however, that reversal portfolios have zero or negative Sharpe ratios after trading cost adjustment. I obtain the same result for reversal portfolios constructed on low-information-asymmetry bonds. However, I show that reversal portfolios on high-asymmetry bonds survive trading cost adjustment.

⁷[Feldhüter and Poulsen \(2018\)](#) also demonstrate that information asymmetry explains only a small percentage of cross-sectional variation in corporate bond bid-ask spreads.

My paper also contributes to the debate on information-driven trading in the corporate bond market. [Asquith et al. \(2013\)](#) analyze the relationship between bond short interest and returns and find no evidence of information-based trading either in investment-grade or in high-yield bonds. [Hendershott, Kozhan, and Raman \(2019\)](#) use similar data on loaned bonds and conclude that information-driven trading is present in high-yield bonds but not in the investment-grade universe. In my paper, high-information-asymmetry bonds are not necessarily high-yield ones. My sample consists mostly of investment-grade bonds, and yet information asymmetry proxies vary a lot in the sample. Therefore, I find evidence of information-based trading in investment-grade bonds. [Han and Zhou \(2014\)](#) also argue that information motives are present in the pricing of bonds of various credit quality by pointing to the positive relationship between microstructure-based information asymmetry measures and bond yield spreads. My paper further emphasizes the circumstances in which information likely stands behind changes in prices of high-asymmetry bonds: when trading volumes are abnormally high, and non-dealer institutions provide liquidity to informed investors.

The latter finding links this paper to the literature on post-crisis liquidity provision in the corporate bond market. The literature has recently documented that liquidity provision has been shifting from dealer banks, which are subject to stricter regulatory requirements, to less constrained bond investors (see, for instance, [Adrian, Boyarchenko, and Shachar 2017](#), [Bessembinder, Jacobsen, Maxwell, and Venkataraman 2018](#), [Choi and Huh 2018](#), and [Dick-Nielsen and Rossi 2018](#)). Dealers still intermediate trading in the latter case but act as pure brokers and do not hold bonds on their books for more than a couple of minutes, avoiding the risk of holding inventory overnight. Despite the emergence of non-dealer liquidity provision, the number of trading days with high customer trade imbalance (substantial changes in dealers' inventory) still exceeds the number of days with sizeable inventory-neutral trading volume in my sample.⁸ Dealers decide on a case-by-case basis whether to let other investors provide liquidity or to accept the inventory risk and provide liquidity themselves. My paper

⁸I consider aggregate dealers' corporate bond inventory in the paper and do not investigate end-of-day inventory changes of individual dealers.

demonstrates that this choice depends on the underlying information asymmetry in the bond, which has not been previously documented in the literature.⁹ I show that dealers tend to pass informed flows to less-informed bond investors and are unlikely to be adversely selected.

The design of my empirical tests follows from a theoretical model of corporate bond trading. In the model, I *assume* that dealers are never adversely selected. An econometrician observing the data generated by the model economy recovers a volume-return relationship (1) and the dependence between volume-return coefficients and information asymmetry that match the ones I find empirically. The methodology of my analysis builds upon [Llorente, Michaely, Saar, and Wang \(2002\)](#). The model I construct extends [Llorente et al. \(2002\)](#) in two dimensions. First, it adapts the asset return dynamics to a defaultable bond rather than a dividend-paying stock. Second, it introduces a noisy market supply representing dealers' inventory.¹⁰ The model falls in a broader class of economies discussed in [Wang \(1994\)](#). The analysis of volume-return relationship also follows the tradition of [Campbell, Grossman, and Wang \(1993\)](#).

Finally, my results contribute to a recent policy debate (see [FINRA 2019](#) proposal). Since late 2004, all corporate bond trades must be reported with a delay of at most 15 minutes. Once reported, trade records become immediately available to all market participants. Some active bond traders have been arguing that there is 'too much' post-trade price transparency in corporate bonds.¹¹ To better study the impact of transparency on liquidity, FINRA proposed a pilot program according to which some bonds become subject to delayed block trade reporting. If the pilot goes through, dealers will be allowed to report big trades in such bonds up to 48 hours later. My paper suggests that this policy change will increase information asymmetry between investors in bonds included in the pilot. Higher asymmetry

⁹[Goldstein and Hotchkiss \(2019\)](#) show that dealers are more reluctant to accept overnight inventory risk in bonds with higher search and inventory costs. Their proxies for the costs associated with OTC market frictions are different from my information asymmetry proxies.

¹⁰[Llorente et al. \(2002\)](#) also regress estimated volume-return coefficients on information asymmetry proxies in the cross-section of stocks to find evidence of information-based trading. They do not distinguish between days with and without changes in aggregate dealers' inventory.

¹¹For liquidity providers, it has become too costly to trade away from large temporary positions every market participant knows about.

is associated with stronger price reversals on days when trading is liquidity-driven. In other words, lower transparency may lead to higher non-fundamental price volatility, which is widely regarded as a negative market feature.

The paper is organized as follows. Section [II](#) talks about the bond sample and the steps I take to estimate a volume-return relationship for individual bonds. Section [III](#) presents estimated volume-return coefficients, and Section [IV](#) investigates its determinants, in particular, information asymmetry proxies, in a cross-section of bonds. Section [V](#) discusses the implications of my results for reversal investment strategies. Section [VI](#) solves a stylized theoretical model of competitive corporate bond trading and discusses a volume-return relationship an econometrician observing such an economy recovers. Section [VII](#) concludes.

II. Data and measurements

A. Data sources

I construct the dataset of corporate bond prices and volumes from Enhanced TRACE tick-by-tick data. The sample is restricted to USD-denominated, fixed-coupon, not asset-backed, non-convertible corporate bonds. I apply the filters of [Dick-Nielsen \(2014\)](#) to clean the TRACE data. I calculate daily corporate bond prices as volume-weighted transaction prices within a given day. Bond characteristics come from Mergent FISD database. I derive the number of mutual funds that own the bond from scraping and processing SEC N-Q forms available through the SEC EDGAR reporting system. The status of the CDS contract on the bond issuer comes from quarterly DTCC Single Name CDS Market Activity reports publicly available at the DTCC website. These reports were machine-read and mentioned entities were matched to the issuers from Mergent FISD dataset. Quarterly DTCC reports are available from Mar 2010, which is the primary reason I start my dataset then; it goes up to Jun 2017. I compute issuer-level characteristics (market capitalization, stock return volatility) using CRSP data. The number of broker-dealers intermediating trades in different

bonds is calculated using the academic version of the TRACE dataset. I talk in more details about the sample in Appendix B.

B. *Sample filtering and ‘active periods’*

I estimate the dynamic volume-return relationship for each bond separately, which requires long enough time-series of returns and volumes for every bond. In a baseline specification of the volume-return relationship (1), I estimate four coefficients in an OLS regression. To avoid over-fitting, I require at least 60 daily observations per bond. However, corporate bonds experience waves of trading activity, as documented in [Ivashchenko and Neklyudov \(2018\)](#). The intervals between trading days with non-zero trading volume might be quite long. Asking for at least 60 consecutive business days is too restrictive, there are very few bonds that satisfy this criterion. Instead, I ask for 60 daily observations where every two successive observations are at most three business days apart.¹²

For some bonds, there is more than one sequence of 60 daily observations where every two consecutive ones are at most three business days apart. I call every such sequence an ‘active period’ and retain all active periods in the sample. I remove all days in between the active periods from the sample. Estimation of the volume-return relationship is carried out per bond per active period.

Also, I remove from the sample all active periods when a bond was either upgraded from high-yield (HY) to investment-grade (IG) territory or downgraded in the opposite direction. [Bao, O’Hara, and Zhou \(2018\)](#) analyze the corporate bond market liquidity around downgrades and find abnormal price and volume patterns associated with insurance companies selling bonds due to regulatory constraints. To ensure that downgrade anomalies do not drive my results, I remove all such periods from my sample. I also remove bonds with less than one year to maturity from the sample. Such bonds are excluded from major bond

¹²Here I follow the methodology of [Bao et al. \(2011\)](#) who study the illiquidity of corporate bonds on the daily data and allow consecutive observations to be at most seven days apart.

market indices, which also drives substantial institutional rebalancing and creates abnormal price patterns that are not the primary focus of this study.

	Mean	Median	S.D.	Min	5th	25th	75th	95th	Max	N.Obs.
Issue size, mln USD	1011.28	750.00	820.94	9.07	166.07	500.00	1250.00	2500.00	15000.00	2720325
Rating	7.73	7.00	3.29	1.00	3.00	6.00	9.00	14.00	21.00	2720325
Age, years	4.15	3.08	3.96	0.00	0.25	1.42	5.75	12.17	31.50	2720325
Maturity, years	8.20	5.58	7.62	1.00	1.42	3.17	9.08	27.33	29.92	2720325
Duration	6.07	4.86	4.24	0.86	1.40	2.94	7.62	15.57	21.57	2720325
Total return, %	0.02	0.02	0.81	-8.19	-1.15	-0.24	0.29	1.18	8.49	2720325
Credit spread, %	2.33	1.70	2.68	0.00	0.59	1.13	2.70	6.01	88.70	2720325
Average bid-ask, %	0.98	0.63	1.02	0.00	0.08	0.29	1.33	3.02	19.99	1550785
No. trades per day	9.06	6.00	12.77	1.00	1.00	3.00	11.00	28.00	2540.00	2720325
No. days since last trade	1.10	1.00	0.35	1.00	1.00	1.00	1.00	2.00	3.00	2718673
C-to-C volume, % of size	0.53	0.02	1.89	0.00	0.00	0.00	0.16	2.83	15.99	2720325
C-to-D volume, % of size	0.01	0.00	3.11	-19.67	-4.00	-0.20	0.32	3.91	17.91	2720325
C-to-D volume , % of size	1.35	0.26	2.81	0.00	0.00	0.06	1.17	6.80	19.67	2720325

Table I. Summary statistics of the filtered bond-day panel. The sample period is from Mar 31, 2010, to Jun 30, 2017. For every bond, I retain only long sequences of daily observations close to each other in the sample. Here, I keep sequences longer than 60 days, where every two daily observations are at most three business days apart. Besides, I exclude from the sample active periods that contain a crossing of the investment-grade/high-yield rating threshold. I keep only bonds with more than one year to maturity in the sample. Size is the amount outstanding. Rating is on a conventional numerical scale from 1 (AAA) to 21 (C). The credit spread is the difference between the observed yield to maturity and yield to maturity of the bond with the same coupons discounted using the Treasury curve as in [Gilchrist and Zakrajšek \(2012\)](#). Average bid-ask spread (realized) is the difference between average client buy and sell prices, expressed as a percentage of the daily average price. It is computed only for the days with at least three trades. C-to-C (client-to-client) trading volume (also, ‘inventory-neutral’ volume) is a minimum between total client purchases and total client sales per bond per day; it is always positive. C-to-D (client-to-dealer) trading volume is the difference between client purchases and client sales; it can be positive (dealers’ inventory decreases) or negative (dealers’ inventory increases) depending on which of the two is greater. The absolute value of the C-to-D trading volume is also the absolute value of the change in aggregate broker-dealer inventory in a given bond. For further details about the sample, see Appendix B. The same summary statistics for a full, unfiltered bond-day panel is in Table C1 in Appendix C.

Table I presents summary statistics of the bond-day panel where only active periods are retained in the sample. My filtered sample includes around 2.7 million bond-day observations that cover approximately 10 thousand distinct active periods between 2010 and 2017 and 5 thousand different bonds issued by 1 thousand unique firms. An average bond in the sample is an investment-grade bond issued about four years ago with approximately eight years left to maturity. Its outstanding notional amount is around 1 billion USD. The average daily

total return of an average bond in the sample is 2 b.p.; the credit spread is approximately 2.3%. The average realized bid-ask spread is about 1%.¹³

C. Volume measures

To construct a proxy for the inventory-neutral trading volume of equation (1), I first compute total daily client purchases from dealers and client sales to dealers; call it V_{it}^{buy} and V_{it}^{sell} respectively for bond i on day t .¹⁴ The minimum of the two is a proxy for inventory-neutral trading volume which I also call ‘C-to-C volume’:

$$\text{Inventory-neutral volume}_{it} = \text{C-to-C volume}_{it} = V_{it}^{(c)} = \min \left\{ V_{it}^{\text{buy}}, V_{it}^{\text{sell}} \right\}$$

It represents trading volume that has no impact on aggregate dealers’ inventory in bond i at the end of the trading day t as compared to day $t-1$; it is non-negative by construction. The difference between client purchases and client sales is a negative change in dealers’ inventory (‘C-to-D volume’):

$$-\text{Change in inventory}_{it} = \text{C-to-D volume}_{it} = V_{it}^{(s)} = V_{it}^{\text{buy}} - V_{it}^{\text{sell}}.$$

The C-to-D volume can be either positive or negative. Positive values represent net purchases by clients from dealers and correspond to a decrease in total broker-dealers’ inventory in bond i on day t . Conversely, negative values of $V^{(s)}$ are increases in dealers’ inventory. When I estimate equation (1), I consider the absolute value of the C-to-D trading volume, $|V_{it}^{(s)}|$. Table I shows that the absolute value of the C-to-D volume is on average several times higher than the C-to-C volume.

Table II demonstrates that there is a positive statistical relationship between the absolute value of changes in inventory and the C-to-C trading volume, but the corresponding correlation coefficient is relatively small. For about two-thirds of bond-active periods, we can not reject the hypothesis that $\text{Corr} \left(V_t^{(c)}, V_t^{(s)} \right) \neq 0$, i.e., bond inventory is equally likely

¹³I present the same summary statistics for the full, unfiltered bond-day panel in Table C1 in Appendix C. Compared to an average bond in the unfiltered sample, the average bond in my sample has a higher outstanding amount, higher credit rating, lower credit spread and bid-ask spread, and lower return.

¹⁴I do not take into account inter-dealer trades when I construct volume proxies.

	Mean	Med.	No.>0	No.<0	No.>0*	No.<0*	No. Obs.
Corr($V_t^{(c)}, V_t^{(s)} $)	0.142	0.130	8356	1466	5052	89	9822
Corr($V_t^{(c)}, V_t^{(s)}$)	-0.052	-0.044	3233	6589	665	2624	9822
Corr($V_t^{(c)}, V_{t-1}^{(c)}$)	0.063	0.028	5758	4064	2920	11	9822
Corr($ V_t^{(s)} , V_{t-1}^{(s)} $)	0.091	0.085	7612	2210	3876	28	9822

Table II. Correlation coefficients between different measures of the trading volume. $V^{(c)}$ is the C-to-C trading volume, $V^{(s)}$ is a signed C-to-D trading volume, and $|V^{(s)}|$ is its absolute value. Each correlation coefficient is estimated per bond per active period. ‘Mean’ and ‘Med.’ are sample average and median values. ‘No. > (<) 0’ is the number of positive (negative) correlation coefficients. ‘No. > (<) 0*’ is the number of positive (negative) coefficients significant at 10% confidence level. The number of observations is the number of bond-active periods.

to fall or to increase on high C-to-C volume days. The persistence of both the C-to-C and the absolute value of the C-to-D trading volume is rather small, as suggested by correlation coefficients in the last two lines of Table II.

D. Proxies for information asymmetry

In empirical tests, I am using several variables to proxy for the extent of information asymmetry between bond investors. Some variables are bond-level proxies:

- the number of mutual funds that hold the bond;
- the number of dealers that intermediate trades in the bond;
- bond outstanding notional amount.

Other variables are issuer-level information asymmetry proxies:

- availability of an active CDS contract on the bond issuer (dummy variable);
- issuer market capitalization;
- realized stock return volatility in an active period when the bond trades actively.

The last two proxies are calculated only for traded companies. Here I assume that informed trading is less likely in bonds that are held by many mutual funds, intermediated by many dealers, have higher outstanding amounts and an actively traded CDS contract on the bond

issuer which is a large firm with lower stock return volatility. Below I justify in more details the use of these variables as the proxies for information asymmetry.

The number of mutual funds that own the bond is related to the number of buy-side analysts scrutinizing bond valuations and the credit quality of the issuer. As in equity literature, I assume that analyst coverage is negatively related to information asymmetry between investors. Similarly, **the number of brokers** intermediating trades in the bond is positively related to sell-side analyst coverage and, hence, negatively related to information asymmetry. The number of active brokers also measures competition among brokers in a given bond. The lack of competition likely affects an average-volume day reversal, β_1 in equation (1), similarly to high information asymmetry: prices of bonds traded in a less competitive market should revert more on average. However, there is no straightforward explanation for why prices for low-competition bonds should revert *less* following high-volume days (the positive relationship between β_2 in equation (1) and information asymmetry) unless low competition among dealers is due to high information asymmetry in the first place.

Issuer and issue sizes are typical proxies for trade informativeness in the literature. Both are related to broader investor base and, again, more in-depth analyst coverage, which supposedly leads to a higher number of investors who are ready to arbitrage out bond misvaluations. As Table III shows, issue and issuer sizes are indeed positively correlated with the numbers of intermediating dealers and mutual funds that own the bond.

	No. funds	Active CDS	Issue size	No. dealers	Issuer size	Stock vol
Active CDS	0.09***					
Issue size	0.59***	0.02				
No. dealers	0.42***	-0.01	0.61***			
Issuer size	0.04***	-0.08***	0.40***	0.30***		
Stock vol	0.04***	-0.10***	-0.13***	0.14***	-0.27***	
Bid-ask	-0.24***	-0.13***	-0.40***	-0.05***	-0.15***	0.41***

Table III. Correlation coefficients between information asymmetry proxies estimated in the cross-section of bonds. If there is more than one active period per bond, the average value across active periods is taken. The total number of bonds (observations) in the sample is 5028. *, **, and *** stand for 10%, 5%, and 1% significance respectively.

The existence of an actively traded CDS contract on the bond issuer is a reasonable proxy for trade informativeness because it is cheaper on average to trade CDS contracts than cash bonds, as [Zawadowski and Oehmke \(2016\)](#) show. Some investors who possess private credit information will rather trade a single-name CDS contract than a bond if the former is available and liquid. Also, the existence of an active CDS contract on the issuer might attract some CDS-bond basis arbitrageurs who trade in the CDS market and the bond market simultaneously. This type of arbitrage does not require any private information about the credit quality of the bond issuer. Hence, an active ‘basis trading’ in some bond implies that only a smaller portion of trading volume in this bond (as compared to an identical bond without an actively traded CDS contract) might be due to private information.

Finally, **stock return volatility** computed for bond issuers over time intervals that constitute the active periods measures uncertainty of bond issuers equity valuations. It is natural to assume that the periods of high uncertainty in equity valuations are also the periods of high asymmetry of information about debt values. Hence, informed trading in equities and bonds might coincide.

I do not use the realized bid-ask spread as an information asymmetry proxy in the paper. It is true that the bid-ask spread might itself be positively related to the extent of informed trading, as in [Glosten and Milgrom \(1985\)](#). However, the mere existence of bid-ask spreads, information or non-information driven, implies price reversals as in [Roll \(1984\)](#), i.e., the ‘bid-ask bounce’ effect. It implies stronger reversals for bonds with wider spreads. Hence, the impact of the bid-ask bounce on the average-day return autocorrelation, β_1 in equation (1), is similar to the expected effect of information asymmetry. The impact of the bid-ask bounce on β_2 and β_3 in equation (1) is unclear because it depends on whether the effect becomes stronger or weaker with higher trading volumes. To avoid these concerns, I use realized bid-ask spreads only as a control variable in my cross-sectional regressions of estimated volume-return coefficients and not as a proxy of informed trading.

III. Volume-return relationship

I estimate equation (1) separately for every bond and every active period rescaling trading volumes such that β_1 measures the first return autocorrelation on average-volume trading days:

$$R_{t+1} = \beta_0 + \beta_1 R_t + \beta_2 R_t \tilde{V}_t^{(c)} + \beta_3 R_t \tilde{V}_t^{(s)} + \epsilon_{t+1}. \quad (2)$$

Above, R_{t+1} is the total bond return between t and $t+1$, $\tilde{V}_t^{(c)}$ is the C-to-C trading volume on day t , standardized¹⁵ for every active period separately, and $\tilde{V}_t^{(s)}$ is *the absolute value* of the C-to-D trading volume (the absolute value of inventory change) on day t , also demeaned and standardized.

On the days when both the C-to-C and the C-to-D trading volumes are at the average level for a given bond in a considered active period, the first return autocorrelation is β_1 . On the days when the C-to-C volume is 1 standard deviation above the mean and the change in inventory is at the average level, the first return autocorrelation is $\beta_1 + \beta_2$. Conversely, when only the C-to-D volume is 1 standard deviation above the average, return autocorrelation equals to $\beta_1 + \beta_3$. Negative values of β_1 would mean that prices revert following average-volume days. Positive values of β_2 and β_3 would mean that prices tend to revert less following high-volume days. In this short section, I present and discuss the estimated volume-return coefficients $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, and in the next section, I investigate in details the relationship between the coefficients and information asymmetry proxies, which is the main focus of this study.

Table IV gives a snapshot of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ estimated for each bond in every active period. The average bond-active period has the first return autocorrelation of approximately -0.33. If the price drops today by 100 b.p. and both trading volumes are at the average level, the price will tend to increase by 33 b.p. tomorrow. One-third of the initial price decrease reverts the next day. The average $\hat{\beta}_2$ of 0.07 suggests that following high C-to-C volume days,

¹⁵De-meaned and divided by the sample standard deviation.

	Mean	Med.	No.>0	No.<0	No.>0*	No.<0*	No. Obs.
$\hat{\beta}_1$	-0.3285	-0.3425	108	9714	0	8761	9822
$\hat{\beta}_2$	0.0716	0.0622	7130	2692	1697	188	9822
$\hat{\beta}_3$	0.0585	0.0568	6928	2894	2054	349	9822
$\hat{\beta}_2 - \hat{\beta}_3$	0.0131	0.0044	5046	4776	3819	3498	9822

Table IV. Summary statistics of the estimated volume-return coefficients of equation (2). Each estimated coefficient is per bond per active period. There are at most nine active periods per bond. Returns are total returns between t and $t + 1$. Trading volumes are demeaned and standardized per bond per active period. Mean and Med. are respectively sample average and sample median. ‘No. $> (<) 0$ ’ is the number of positive (negative) coefficients. ‘No. $> (<) 0^*$ ’ is the number of positive (negative) coefficients significant at 10% confidence level. The number of observations is the number of bond-active periods.

prices tend to revert less. In a previous example, if the initial 100 b.p. price decrease was accompanied by 1 standard deviation above-average C-to-C trading volume, then the next day reversal would be close to one-fourth rather than one-third. The average $\hat{\beta}_3$ of around 0.06 suggests that prices revert comparably less following high C-to-D volume days either. Both $\hat{\beta}_2$ and $\hat{\beta}_3$ are predominantly positive, and the difference between the two is equally likely to be positive or negative.

At this stage, we can not infer much from estimated volume-return coefficients. The signs and the magnitudes of the coefficients certainly look reasonable. Strongly negative $\hat{\beta}_1$ is a reflection of high illiquidity of the corporate bond market. The values of $\hat{\beta}_2$ and $\hat{\beta}_3$ are close; hence, both types of trading volume interact statistically similarly with reversals. Positive $\hat{\beta}_2$ and $\hat{\beta}_3$ can be consistent with the presence of informed trading, but can also reflect correlated trading volumes, or the interaction of the bid-ask bounce or bond riskiness with the trading volume. In the next section, I investigate explanatory factors of the cross-sections of volume-return coefficients with a particular focus on the impact of information asymmetry.

IV. Determinants of volume-return coefficients

A. Empirical design

In the introduction, I put forward an intuition on how volume-return coefficients β_1 , β_2 , and β_3 in equation (2) should vary with information asymmetry. In particular, I suggest that more information asymmetry implies lower β_1 (stronger reversals on average), higher β_2 (weaker reversals following high-volume days when dealers' inventory does not change much), and no particular effect on β_3 (no difference in reversals between high- and low-information-asymmetry bonds following days when dealers' inventory changes a lot). One gets the same relationship between volume-return coefficients in a theoretical model a-là [Llorente et al. \(2002\)](#) extended with noisy changes in market supply (dealers' inventory) that are independent from the arrival of private news. I present such a model formally in Section VI. In this section, I am testing the predictions of the model empirically in the cross-section of bonds.

The estimates $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ obtained in the previous section are per bond and per active period. There is more than one active period for about every fifth bond in the sample, but there are at most nine active periods per bond. I take bond averages to obtain the cross-section of coefficients, and in the rest of this section, I fit explanatory linear models to this cross-section.¹⁶ Call $\hat{\beta}_{n,i}$ a column-vector of estimates ($n = 1$, 2, or 3 and $i \in \{1, \dots, N\}$ where N is the total number of bonds). I fit the following model for each n (i.e., each

¹⁶active periods are asynchronous across bonds. Hence, one needs to make additional assumptions to investigate the co-movement of volume-return coefficients. I attributed the estimated coefficients to quarters in the proportion of the active period time in a given quarter and extracted time fixed effects from the bond-quarter panel to find that there is virtually no common time variation in the coefficients (unreported).

volume-return coefficient) separately:

$$\hat{\beta}_{n,i} = c_{n,1} \underbrace{(\text{No. funds, CDS, Issue/issuer size, No. dealers, -Equity volatility})_i}_{\text{Info asymmetry proxies}} + \\ + c_{n,2} \underbrace{(\text{Bid-ask, C-to-C/D volume correlation, Bond volatility, Credit spread})_i}_{\text{Controls}} + \\ + c_{n,0} + \epsilon_{n,i}, \quad (3)$$

where $c_{n,1} \in \mathbb{R}^6$, $c_{n,2} \in \mathbb{R}^5$, and $\epsilon_{n,i}$ for every n is an i.i.d. zero-mean Normal. If my intuition about the dependence of volume-return coefficients on information asymmetry proxies is correct, I should find $c_{1,1} > 0$, $c_{2,1} < 0$, and $c_{3,1} = 0$.

I include five controls in the baseline specification (3): realized bid-ask spread, the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(S)}$, realized bond return volatility, and the credit spread. Volume autocorrelations and return volatility are estimated per bond per active period, and then bond averages are computed if there is more than one active period per bond.

The realized bid-ask spread controls for bond illiquidity.¹⁷ Wider spreads are associated with more illiquid bonds that tend to have stronger price reversals even if the information is symmetric, buy and sell orders arrive randomly, and the fundamental value of the security never changes (the ‘bid-ask bounce’ effect of [Roll 1984](#)). In principle, bid-ask spreads also widen with the asymmetry of information, as in [Glosten and Milgrom \(1985\)](#), and that is why the literature often uses bid-ask spreads as a measure of information asymmetry. I do not do so because multiple non-informational reasons might explain different bid-ask spreads in the cross-section of bonds, for instance, competition between dealers, different inventory holding costs, or counterparty search costs. The bid-ask spread as the illiquidity control is the most relevant for the regressions of $\hat{\beta}_1$.

Volume correlations control for the persistence of trade flow and price impact. Recall from Table [IV](#) that returns tend to continue following high C-to-C and C-to-D volume days

¹⁷[Schestag, Schuster, and Uhrig-Homburg \(2016\)](#) provide a detailed comparison of different bond illiquidity measures. In light of their results on different measures one can compute using tick-by-tick TRACE data, the realized bid-ask spread looks like a reasonable choice for this paper. I obtained similar results with alternative bond illiquidity measures as well (Amihud, Roll, price inter-quartile range).

(positive $\hat{\beta}_2$ and $\hat{\beta}_3$). I want to link it with the presence of informed trading, but one would find the same signs of volume-return coefficients if trade flows were persistent. Imagine that some investor executes a big buy trade over two business days.¹⁸ On each day, her trades have a price impact, and returns tend to continue (or revert less). So, correlated volumes would generate the relationship between volumes and future returns similar to one of the asymmetric information and returns. I control for this alternative explanation by including the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$ in the group of control variables. These controls are the most relevant for the regressions of $\hat{\beta}_2$ and $\hat{\beta}_3$.

The next control is the average realized bond return volatility. Riskier bonds tend to experience larger price swings, even if underlying risks are not directly related to information asymmetry. In the cross-section, some bonds are riskier than the other, and it might explain some differences between estimated volume-return coefficients. Same happens in the theoretical model of Section VI. The desired dependence of volume-return coefficients on information asymmetry is obtained when holding unconditional bond return variance fixed. To mimic this condition in the empirical analysis, I include realized bond return volatility as a control variable in every regression. It is relevant for the regressions of all three volume-return coefficients. I further include average credit spread as a control variable to make sure that I compare bonds with the same riskiness. One can easily find a high-yield and an investment-grade bond with comparable levels of return volatility in some periods, but their credit spreads must be different.

Table V presents summary statistics of the cross-section of estimated volume-return coefficients to be explained, information asymmetry proxies, and control variables. The average bond in the cross-section is owned by 35 mutual funds and about the same number of dealers intermediate trades in this bond. The bond is issued by a large company (76 bln USD market cap) and has an outstanding notional amount of around 800 mln USD. The average realized

¹⁸This hypothesis may not be very realistic since on the corporate bond market one may get better execution prices trading higher volumes as shown in [Edwards, Harris, and Piwowar \(2007\)](#). This may also explain why the average autocorrelation of $\tilde{V}_t^{(c)}$ is relatively low in the data (see Table II).

	Mean	Median	S.D.	Min	5th	25th	75th	95th	Max	N.Obs.
$\hat{\beta}_1$	-0.31	-0.33	0.12	-0.62	-0.48	-0.40	-0.24	-0.09	0.05	5028
$\hat{\beta}_2$	0.07	0.06	0.12	-0.48	-0.10	0.01	0.12	0.25	0.79	5028
$\hat{\beta}_3$	0.06	0.06	0.10	-0.33	-0.10	-0.00	0.11	0.21	0.49	5028
No. mutual fund owners	35.47	28.41	31.31	0.00	0.00	12.91	49.55	97.29	230.46	5028
Active CDS (dummy)	0.44	0.00	0.50	0.00	0.00	0.00	1.00	1.00	1.00	5028
Issue size, bln USD	0.82	0.60	0.70	0.01	0.07	0.40	1.00	2.25	9.39	5028
No. dealers	33.98	29.50	15.13	7.96	17.65	23.96	39.89	65.46	168.72	5026
Issuer size, bln USD	76.09	40.92	92.71	0.02	2.58	13.44	115.85	236.12	761.79	4693
Stock return volatility, %	1.77	1.57	0.84	0.65	0.93	1.23	2.06	3.25	10.52	4683
Average bid-ask, %	1.05	0.77	0.83	0.07	0.22	0.46	1.38	2.82	8.66	5028
C-to-C volume correlation	0.08	0.06	0.11	-0.18	-0.05	-0.00	0.14	0.29	0.66	5028
C-to-D volume correlation	0.10	0.10	0.09	-0.24	-0.05	0.04	0.15	0.25	0.79	5028
Bond return volatility, %	0.72	0.59	0.51	0.05	0.17	0.36	0.94	1.68	4.96	5028
Credit spread, %	2.42	1.74	2.85	0.14	0.58	1.11	2.78	6.39	68.96	5028

Table V. Summary statistics of the cross-section of volume-return coefficients and their predictors. The sample contains bond averages computed across all active periods in case there is more than one for a given bond. The number of fund owners on a given trading date represents the number of mutual funds that claim to own a bond as of the latest available SEC N-Q form filing. ‘Active CDS’ is a dummy variable that equals 1 for all bonds of the issuer on all days in a given quarter if the CDS on this issuer is in a list of top thousand actively traded single-name CDS contracts in that quarter according to DTCC. ‘Issue size’ is an outstanding notional amount of a bond issue, ‘issuer size’ is the market capitalization of an issuer (if a traded company). The number of dealers is the number of broker-dealers that intermediate trades in a bond on each trading day. Stock return volatility is the realized volatility in a given active period for a given issuer. For further details, see Appendix B.

bid-ask spread of the bond is 105 b.p., and its credit spread is 242 b.p. 44% of bonds in the sample have an actively traded CDS on the bond issuer. There is substantial variation in both the left-hand side and right-hand side variables of regressions (3) as Table V shows.

B. Main results

Tables VI–VIII present estimated regressions (3) of volume-return coefficients on information asymmetry proxies and controls. Table VI contains the results for $\hat{\beta}_1$. Observe that the number of fund owners, the CDS dummy, issue and issuer size, and the number of intermediating dealers, all have a significantly positive impact on $\hat{\beta}_1$ if included in the regression separately. In joint models 7 (all bonds) and 8 (bonds issued by traded firms only), the loading on the CDS dummy becomes insignificant but on the negative stock return volatil-

ity – significantly positive. These results suggest that average-day price reversals become more pronounced ($\hat{\beta}_1$ becomes more negative) for higher information asymmetry bonds: the bonds with fewer fund owners and intermediating dealers, no actively traded CDS contract on the issuer, lower issue and issuer size, and high stock return volatility. Observe also in Table VI that the coefficient on the average bid-ask spread is significant with a reasonable sign. Higher bid-asks are associated with stronger reversals.

Interestingly, in Table VI, C-to-C and C-to-D volume persistence both enter the models for $\hat{\beta}_1$ significantly but with different signs. Following an average-volume day, higher C-to-C volume persistence implies less strong reversals, while higher C-to-D volume persistence implies stronger reversals holding other bond characteristics equal. One can interpret this finding as follows: if an investor has to trade persistently high volumes over several consecutive days with a dealer hence asking the dealer for immediacy, trading costs in such trading arrangement will be higher than when another bond investor supplies liquidity.

The link between high information asymmetry and strong price reversals following average-volume days relates to a recent policy debate on delayed corporate bond trade dissemination. Now, dealers must report corporate bond trades to TRACE at most 15 minutes after trade execution. A pilot program, currently under discussion, proposes a 48 hours delay between trade execution and reporting for some bonds (see [FINRA 2019](#)). From the perspective of the results presented in Table VI, such policy change might lead to stronger price reversals in bonds selected for the pilot because the policy increases information asymmetry between investors. Since we talk about average-volume days here, trading on such days is primarily liquidity-driven and stronger reversals can be interpreted as higher non-fundamental price volatility (bond valuations do not change when prices do not reveal any fundamental information). Higher volatility unrelated to fundamentals is a likely (and negative) consequence of the delayed trade dissemination pilot if it goes through.

Table VII presents the results for $\hat{\beta}_2$. Recall that higher β_2 means less strong reversals following days when investors trade a lot essentially with each other and dealers do not

	Dependent variable: $\hat{\beta}_1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	-0.331*** (0.004)	-0.290*** (0.004)	-0.385*** (0.005)	-0.391*** (0.006)	-0.309*** (0.004)	-0.283*** (0.005)	-0.408*** (0.006)	-0.409*** (0.007)
Average bid-ask	-0.067*** (0.005)	-0.077*** (0.005)	-0.067*** (0.005)	-0.118*** (0.006)	-0.089*** (0.005)	-0.084*** (0.005)	-0.079*** (0.006)	-0.092*** (0.006)
C-to-C vlm corr.	0.253*** (0.016)	0.327*** (0.015)	0.167*** (0.016)	0.187*** (0.016)	0.308*** (0.016)	0.338*** (0.016)	0.140*** (0.016)	0.143*** (0.016)
C-to-D vlm corr.	-0.166*** (0.015)	-0.198*** (0.014)	-0.060*** (0.014)	-0.161*** (0.014)	-0.189*** (0.015)	-0.217*** (0.015)	-0.076*** (0.015)	-0.087*** (0.015)
No. funds	0.001*** (0.0001)						0.0004*** (0.0001)	0.0003*** (0.0001)
CDS dummy		0.006** (0.003)					0.003 (0.003)	0.002 (0.003)
Issue size			0.079*** (0.004)				0.058*** (0.005)	0.052*** (0.005)
No. dealers				0.003*** (0.0001)			0.001*** (0.0002)	0.001*** (0.0002)
Issuer size					0.0002*** (0.00002)			0.0001*** (0.00002)
-Equity volatility						0.355 (0.288)	0.494* (0.275)	
Risk controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,028	5,028	5,028	5,026	4,693	4,683	5,026	4,681
R ²	0.310	0.249	0.380	0.330	0.278	0.259	0.392	0.405

Note:

*p<0.1; **p<0.05; ***p<0.01

Table VI. Cross-sectional regressions of $\hat{\beta}_1$. Each model is an OLS regression with heteroscedasticity-consistent standard errors. $\hat{\beta}_1$ is averaged for every bond across all active periods, so are the predictors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. ‘No. funds’ is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. ‘No. dealers’ is the average number of unique dealers that intermediate trades in each bond. Equity volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

hold any additional inventory by the end of the trading day. I expect $\hat{\beta}_2$ to be increasing in information asymmetry: reversals must be less strong for high asymmetry bonds when informed trading is most likely, i.e., after high C-to-C volume days. Observe first in Table VII that all information asymmetry proxies enter the models for $\hat{\beta}_2$ significantly when included separately (models 1 to 6) except for stock return volatility. The signs of all asymmetry proxies are as expected: higher information asymmetry implies higher $\hat{\beta}_2$. In a joint model 7 (bonds issued by public and private firms) the CDS dummy turns insignificant while in a joint model 8 (bonds issued by public firms only) the issuer size becomes insignificant and flips a sign. Otherwise, a joint model 8 says that bonds with fewer mutual fund owners and

	Dependent variable: $\hat{\beta}_2$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.088*** (0.005)	0.075*** (0.005)	0.101*** (0.006)	0.113*** (0.007)	0.074*** (0.005)	0.067*** (0.006)	0.119*** (0.008)	0.112*** (0.009)
Average bid-ask	0.001 (0.005)	0.004 (0.005)	0.002 (0.005)	0.021*** (0.005)	0.009* (0.005)	0.007 (0.005)	0.011* (0.006)	0.013** (0.006)
C-to-C vlm corr.	0.070*** (0.017)	0.041** (0.016)	0.089*** (0.017)	0.096*** (0.018)	0.046*** (0.016)	0.040** (0.016)	0.106*** (0.018)	0.105*** (0.017)
C-to-D vlm corr.	-0.104*** (0.018)	-0.090*** (0.018)	-0.132*** (0.019)	-0.106*** (0.018)	-0.097*** (0.018)	-0.094*** (0.018)	-0.120*** (0.020)	-0.127*** (0.020)
No. funds	-0.0004*** (0.00005)						-0.0002*** (0.0001)	-0.0002*** (0.0001)
CDS dummy		-0.007** (0.003)					-0.005 (0.003)	-0.006* (0.003)
Issue size			-0.023*** (0.003)				-0.009** (0.004)	-0.012*** (0.004)
No. dealers				-0.001*** (0.0001)			-0.001*** (0.0002)	-0.001*** (0.0002)
Issuer size					-0.00003* (0.00002)		0.00002 (0.00002)	
-Equity volatility						-0.352 (0.304)		-0.575* (0.312)
Risk controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,028	5,028	5,028	5,026	4,693	4,683	5,026	4,681
R ²	0.021	0.012	0.024	0.025	0.013	0.013	0.030	0.034

Note: *p<0.1; **p<0.05; ***p<0.01

Table VII. Cross-sectional regressions of $\hat{\beta}_2$. Each model is an OLS regression with heteroscedasticity-consistent standard errors. $\hat{\beta}_2$ is averaged for every bond across all active periods, so are the predictors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. ‘No. funds’ is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. ‘No. dealers’ is the average number of unique dealers that intermediate trades in each bond. Equity volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

intermediating dealers, lower outstanding amounts, no actively traded CDS contract, and higher stock return volatility exhibit less strong price reversals following high C-to-C volume days.

Also, observe in Table VII that the loading on the C-to-C volume persistence is positive and significant. It means that if high C-to-C volumes are positively correlated over time, reversals will be less strong due to a repetitive price impact. The bid-ask spread enters joint models of Table VII with significantly positive coefficients: the bonds with higher bid-ask spreads tend to revert less following high C-to-C volume days. If I treated the bid-ask spread

	Dependent variable: $\hat{\beta}_3$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.070*** (0.004)	0.074*** (0.004)	0.086*** (0.005)	0.091*** (0.006)	0.089*** (0.004)	0.076*** (0.005)	0.088*** (0.006)	0.097*** (0.007)
Average bid-ask	-0.057*** (0.004)	-0.058*** (0.004)	-0.060*** (0.004)	-0.053*** (0.004)	-0.054*** (0.004)	-0.057*** (0.004)	-0.051*** (0.005)	-0.049*** (0.005)
C-to-C vlm corr.	-0.017 (0.014)	-0.006 (0.014)	0.009 (0.015)	0.012 (0.015)	0.005 (0.014)	-0.015 (0.014)	0.011 (0.015)	0.008 (0.016)
C-to-D vlm corr.	0.007 (0.015)	0.001 (0.015)	-0.010 (0.016)	-0.003 (0.015)	-0.016 (0.016)	0.002 (0.015)	-0.012 (0.016)	-0.021 (0.017)
No. funds	0.0002*** (0.00005)						0.0003*** (0.0001)	0.0002*** (0.0001)
CDS dummy		0.005* (0.003)					0.005* (0.003)	0.002 (0.003)
Issue size			-0.007*** (0.002)				-0.010*** (0.003)	-0.005 (0.004)
No. dealers				-0.0004*** (0.0001)			-0.0004** (0.0002)	-0.0002 (0.0002)
Issuer size					-0.0001*** (0.00002)			-0.0001*** (0.00002)
-Equity volatility						0.020 (0.284)	0.426 (0.280)	
Risk controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,028	5,028	5,028	5,026	4,693	4,683	5,026	4,681
R ²	0.083	0.082	0.083	0.085	0.083	0.070	0.091	0.087

Note: *p<0.1; **p<0.05; ***p<0.01

Table VIII. Cross-sectional regressions of $\hat{\beta}_3$. Each model is an OLS regression with heteroscedasticity-consistent standard errors. $\hat{\beta}_3$ is averaged for every bond across all active periods, so are the predictors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. ‘No. funds’ is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. ‘No. dealers’ is the average number of unique dealers that intermediate trades in each bond. Equity volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

as a proxy for information asymmetry, this sign on the bid-ask would have been in line with the signs on other information asymmetry proxies.

Table VIII presents the regressions for $\hat{\beta}_3$. The interpretation of β_3 is analogous to β_2 , but now we are talking about the reversals following days when dealers’ inventory changes a lot. Higher β_3 means that prices tend to revert less following high C-to-D volume days. Unlike for β_2 , I do not expect to find any particular dependence of β_3 on information asymmetry because dealers would rather pass high-asymmetry bonds to other investors and would not hold excess inventory in bonds with less transparent valuations.

Table [VIII](#) shows that there is indeed no clear-cut dependence of $\hat{\beta}_3$ on information asymmetry. For instance, the number of mutual fund bond owners and the CDS dummy have significantly positive loadings in models 1 and 2 (opposite to what information asymmetry explanation predicts), while issuer and issue size and the number of dealers have significantly positive loadings in models 3–5 (in line with information asymmetry explanation). In joint models 7 and 8 as well, there are both positive and negative loadings on the variables of interest. In particular, in model 8, only the number of mutual fund bond owners and issuer size have significant loadings, but they are of opposite signs.

Tables [VI](#)–[VIII](#) show that high-information-asymmetry bonds experience on average stronger price reversals than low asymmetry bonds. However, following high C-to-C trading volume days, this ‘gap’ in reversals closes; such thing does not happen following days with high C-to-D trading volume. How large is this difference in reversals between high and low asymmetry bonds? To answer this question, I take the last models from Tables [VI](#)–[VIII](#) (models number 8) and compute average values of volume-return coefficients predicted by fitted models for different deciles of information asymmetry proxies.¹⁹ The bonds with the most information asymmetry are in the first decile for every proxy except for stock return volatility (here, the most asymmetry is in the tenth decile). Conversely, the bonds with the least information asymmetry are in top deciles (bottom decile of stock return volatility). I keep control variables fixed at the median level to ensure that predicted values of volume-return coefficients vary only due to changing information asymmetry.

Figure 3 presents the results. The left panel shows the average values of $\hat{\beta}_1$. They are decreasing monotonically from -0.2 for the bonds with little or none information asymmetry to almost -0.4 for the bonds with the highest asymmetry. The predicted reversal for high-asymmetry bonds is almost twice stronger than for low-asymmetry bonds following average-volume days. The middle panel in Figure 3 shows an additional impact of high C-to-C volumes on next-day reversals. The average values of $\hat{\beta}_2$ are monotonically increasing

¹⁹The results look almost identical when I use models number 7 for public and non-public firms with all proxies included (unreported).

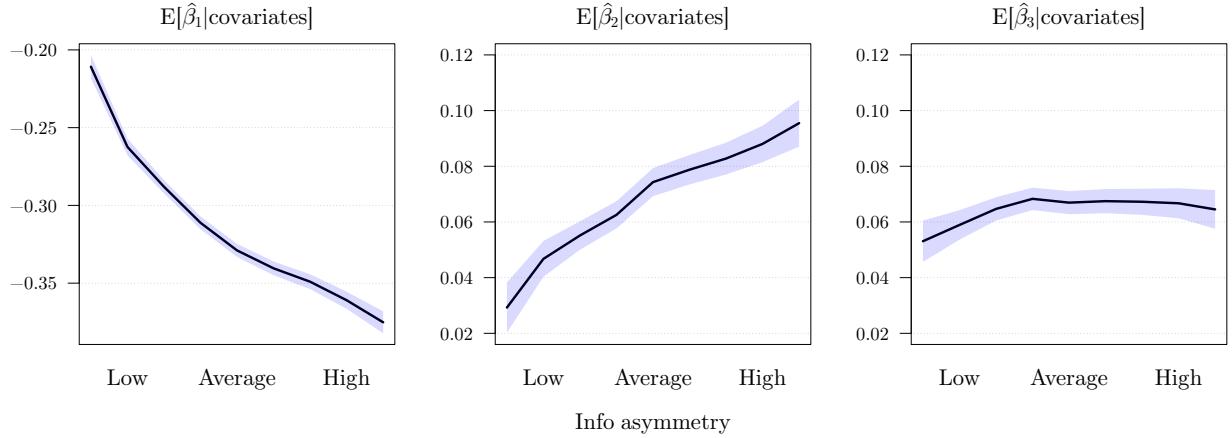


Figure 3. Point estimates and confidence intervals for the expected values of volume-return coefficients. The calculations are based on models (8) from Tables VI–VIII. On the x-axes from left to right are the deciles of information asymmetry proxies. For instance, ‘Low asymmetry’ bond is the one that has the number of fund owners, active CDS dummy, issue size, number of dealers, and issuer size all in the 90th percentile and stock volatility in the 10th percentile. ‘High asymmetry’ bond has the number of fund owners, active CDS dummy, issue size, number of dealers, and issuer size all in the 10th percentile and stock volatility in the 90th percentile. All other covariates from the regression models (average bid-ask spread, volume correlations, return volatility, and credit spread) are fixed at the median level. Solid lines are points estimates and shaded areas around them are 95% confidence bands.

from 0.02 for low-asymmetry to 0.10 for high-asymmetry bonds. It means that every additional standard deviation of the C-to-C volume reduces the difference in next-day reversals between high- and low-asymmetry bonds by almost 0.08. Figure C1 in Appendix C shows that following a day with the C-to-C trading volume 2 standard deviations above the average, there is practically no difference in reversals between high- and low-asymmetry bonds. Finally, the right panel in Figure 3 demonstrates that predicted $\hat{\beta}_3$ is relatively insensitive to the degree of information asymmetry; the average $\hat{\beta}_3$ stays close to 0.06 as information asymmetry varies. This result implies that the average difference in reversals between high- and low-asymmetry bonds stays the same following days when dealers’ inventory changes a lot. The evidence presented in Figure 3 suggests that information-driven trading in corporate bonds exists, and it is much more likely when investors essentially trade with each other within one trading day rather than when they trade with dealers.

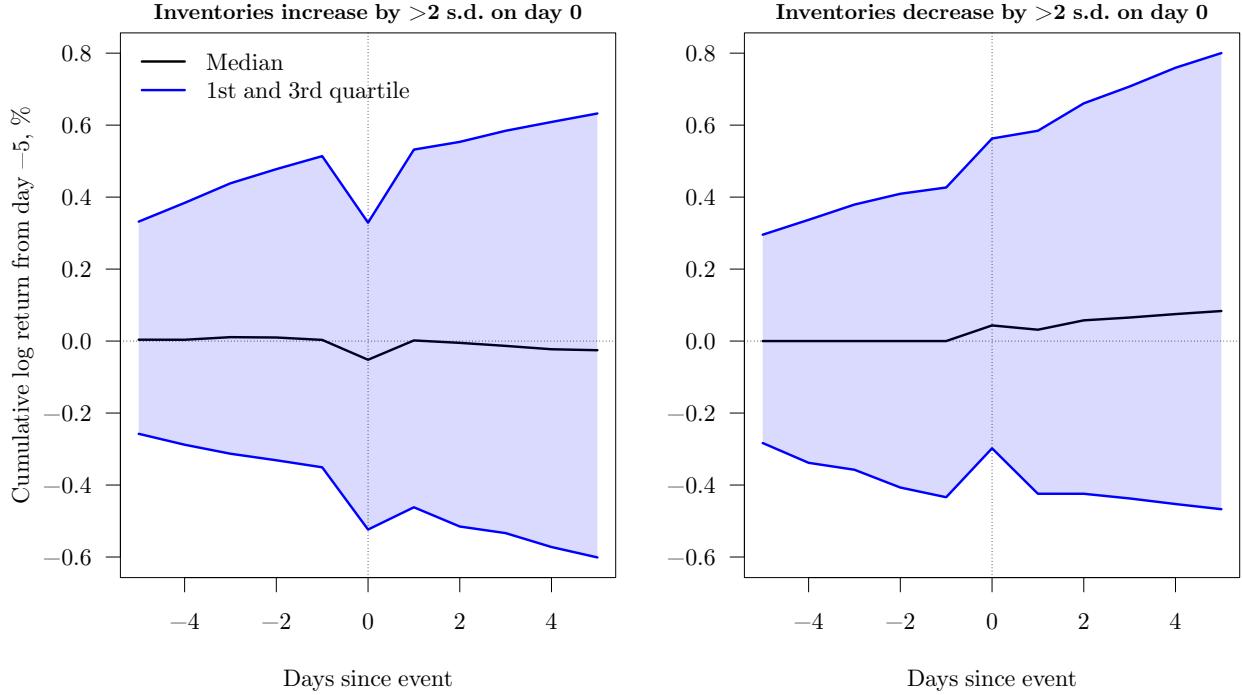


Figure 4. Cumulative returns around days with large bond inventory changes. The ‘event’ that happens on day 0: broker-dealers bond inventory increases or decreases by more than 2 standard deviations (computed per bond per active period) *and* it is the only type of trading that occurs on day 0 (inventory stays on the books till day 1). Daily log price returns are cumulated from day -5. Returns are computed using clean prices and do not contain accrued interest.

To provide additional evidence that dealers are very unlikely to be adversely selected (to trade with a privately informed investor) in the corporate bond market, I plot typical cumulative return paths around days when dealers’ inventory changes a lot only in one particular direction. In terms of two types of volume introduced in Section II, such days correspond to high C-to-D volume and *zero* C-to-C volume. Figure 4 plots the results of such ‘event study’. On the left panel, a more interesting one, dealers’ inventory increase by at least 2 standard deviations (per bond per active period) on day 0. In other words, on day 0, investors sell a lot of bonds to dealers hence asking for immediacy. There is a well-pronounced drop in cumulative returns on day 0 regardless of whether prices were going up or down before the event. Cumulative returns rebound to their pre-event paths on day 1. It means that additional inventory that dealers acquired on day 0 is sold (at least partially)

on day 1 at higher prices. Even for the worst-performing bonds, dealers could sell at higher prices 2-3 days after the initial increase in inventory. The right panel of Figure 4 presents similar cumulative return patterns for the days when dealers' inventory reduces by more than 2 standard deviations (some investors are willing to buy a lot of bonds and do not want to wait for a selling investor to come to the market). There is a pronounced spike in cumulative returns on day 0. On day 1, prices are lower than on day 0 except for the cases when bonds have been performing well pre-event. Such a situation (dealers sell short-term ‘winners’) is the only case in Figure 4 when prices do not move in dealers’ favor post-event. In all other cases, dealers benefit from price movements on and right after the event day, which is consistent with a finding that dealers are unlikely to trade with an informed counterparty.

C. Further evidence

There are firms that have many bonds outstanding. These bonds may differ in coupon rates, maturity, embedded options, and other characteristics. I investigate how volume-return coefficients differ across bonds of the same issuer. In Table IX, I present the estimates of model (3) only for firms with more than fifteen bonds outstanding. I include issuer fixed effects in the regression models; such fixed effects represent the average values of volume-return coefficients for different issuers. Thus, Table IX shows within-firm dependence of volume-return coefficients on information asymmetry. I find that the impact of information asymmetry on $\hat{\beta}_1$ and $\hat{\beta}_2$ (and the lack of impact on $\hat{\beta}_3$) holds for the bonds of the same issuer. It suggests that private information some investors might possess is not only issuer-level (which is most likely private news about the credit quality of the issuer) but also bond-level. The bond-level information can be, for instance, private knowledge about liquidity trades of other investors, which yields a better estimate of price pressures and subsequent price reversals.²⁰ It can also be private knowledge about the exercise probability of embedded

²⁰I remain agnostic about a mechanism through which some investors may learn valuable information about price pressures. [Barbon, Di Maggio, Franzoni, and Landier \(2018\)](#) suggest that there is information leakage from brokers to clients in the equity market.

options. Most bonds in my sample are callable; issuers have a right to redeem them at pre-specified dates before maturity. An early call changes the duration of a bond and, therefore, its risk profile. Superior knowledge about the likelihood of an early call gives advantage in predicting bond returns prior to call announcements.

	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_3$
Average bid-ask	-0.078*** (0.012)	-0.090*** (0.012)	0.006 (0.013)	0.008 (0.012)	-0.007 (0.010)	-0.005 (0.009)
No. funds	0.0004*** (0.0001)	0.0003*** (0.0001)	-0.0002** (0.0001)	-0.0002* (0.0001)	0.0002 (0.0001)	0.0002 (0.0001)
CDS dummy	0.022 (0.019)	-0.009 (0.017)	0.001 (0.020)	-0.004 (0.020)	0.001 (0.016)	-0.005 (0.015)
Issue size	0.038*** (0.007)	0.032*** (0.007)	-0.001 (0.005)	-0.003 (0.005)	-0.005 (0.005)	-0.005 (0.005)
No. dealers	0.001*** (0.0003)	0.002*** (0.0003)	-0.001** (0.0003)	-0.001** (0.0003)	-0.0005* (0.0003)	-0.0004 (0.0003)
Issuer size		0.001*** (0.0001)		0.0001 (0.0001)		-0.0002 (0.0001)
-Equity volatility		3.028*** (0.784)		-2.477*** (0.937)		1.214 (0.839)
Issuer FE	YES	YES	YES	YES	YES	YES
Risk controls	YES	YES	YES	YES	YES	YES
Vlm correlations	YES	YES	YES	YES	YES	YES
Observations	1,927	1,837	1,927	1,837	1,927	1,837
R ²	0.555	0.566	0.116	0.134	0.228	0.213

Note:

*p<0.1; **p<0.05; ***p<0.01

Table IX. Cross-sectional regressions of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ for large issuers only. Each model is an OLS regression with heteroscedasticity-consistent standard errors. Volume-return coefficients are averaged for every bond across all active periods, so are the predictors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. ‘No. funds’ is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. ‘No. dealers’ is the average number of unique dealers that intermediate trades in each bond. Equity volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

In Appendix C, I present further empirical results. Table C2 estimates equation (3) for investment-grade (IG) and high-yield (HY) subsamples separately. The markets for IG and HY bonds have different institutional clientele because of regulatory restrictions, but information asymmetry proxies I use should work within each subsample. Table C2 confirms that it is indeed the case for $\hat{\beta}_1$ and $\hat{\beta}_2$. $\hat{\beta}_1$ tends to decrease and $\hat{\beta}_2$ to increase

with information asymmetry both for IG and HY bonds. In the regressions for $\hat{\beta}_3$, there are fewer significant coefficients compared to the regressions of $\hat{\beta}_1$ and $\hat{\beta}_2$, and the signs of the coefficients are inconclusive about the impact of information asymmetry on reversals following high C-to-D volume days. Hence, Table C2 confirms the results that have been established in the pooled sample.

I also consider alternative specifications of equation (2) to address the omitted variable problem that may render the estimates of volume-return coefficients biased. In Appendix C, I present key results for volume-return coefficients estimated controlling for either market returns or trading volumes (included as linear terms in addition to the interactions with returns) in equation (2). Tables C3 and C5 present summary statistics of volume-return coefficients for these two cases, while Tables C4 and C6 show the dependence on information asymmetry proxies. Figures C2 and C3, the counterparts of Figure 3, demonstrate how predicted volume-return coefficients vary with information asymmetry. Clearly, the main result of the empirical analysis remains intact. $\hat{\beta}_1$ decreases as information asymmetry grows while $\hat{\beta}_2$ increases; the impact of asymmetry on $\hat{\beta}_3$ is neutral.

V. Implications for investment strategies

Corporate bond price reversals depend on the extent of information asymmetry in a given bond, as my empirical analysis shows. What does it imply for the design of the short-term corporate bond reversal strategy? In this section, I show that the reversal strategy earns more if information asymmetry is taken into account in portfolio formation.

I start by constructing reversal portfolios as in [Bai et al. \(2019\)](#). At every rebalancing date (which is monthly) bonds are double sorted on previous month's credit rating and return. In [Bai et al. \(2019\)](#) each sorting is into quintiles but since my sample is smaller I sort into rating terciles and return quintiles, a total of 15 bins. I only consider the long part of the reversal portfolio: this is a simple average of size-weighted returns in the top reversal

quintile (lowest past returns) across three rating terciles.²¹ The rebalancing is at the end of each month. I consider an unfiltered bond-month sample, i.e., I do not restrict the sample to active periods and do not remove the crossing of IG/HY threshold (I would introduce a look-ahead bias if I did so). I do require the bonds to have, as of the sorting date, an outstanding amount of at least 200 mln USD and a 12-month average of the realized bid-ask spread of at most 100 b.p. The latter helps to bring down the transaction cost of the reversal strategy which is usually very high due to high portfolio turnover. I use the 12-month average of the realized bid-ask spread to account for transaction costs. I also extend the sample back to 2005 to compare the performance of the reversal strategy pre- and post-2008 crisis.

In addition to a long-reversal portfolio, I consider its two sub-portfolios separately. The first sub-portfolio contains the bonds with a below-median number of mutual fund bond-holders as of the sorting date.²² This sub-portfolio contains bonds with supposedly more information asymmetry. The second sub-portfolio contains the bonds with an above-median number of mutual fund bondholders (less information asymmetry). The results of the previous section suggest that in-sample and following average-volume periods the reversals are stronger for bonds with more information asymmetry. So, one might expect the reversal portfolio with more information asymmetry to outperform the reversal portfolio with less information asymmetry out-of-sample .

Table X presents performance measures of three reversal portfolios in comparison to the market portfolio. Between Oct 2005 and Jun 2017 average long-reversal portfolio returns unadjusted for trading costs were around 8.4% per year. The sub-portfolio with many fund owners earned around 8% while the portfolio with few fund owners earned around 9%, which is 4.5 times more than the market portfolio. The volatility of the sub-portfolio with few fund owners was also lower which translates into a superior risk-adjusted performance

²¹I do not consider a short leg here for two reasons. First, in the sample I work with shorting top-performing corporate bonds was not profitable. Second, I do not have reliable estimates for the cost of shorting.

²²For sorting, I take the variable ‘number of mutual fund owners’ as before but with a lag of 6 months. Since N-Q forms are reported semiannually, it ensures that I am not sorting on the information not yet available at the sorting date.

	Cum trading costs				Net trading costs			
	Mean	S.D.	SR	IR	Mean	S.D.	SR	IR
Long reversal (LR)	8.40	6.44	1.12	1.83	1.96	6.34	0.13	0.18
LR: many funds	8.02	7.09	0.97	1.40	1.39	6.99	0.04	0.01
LR: few funds	9.01	6.11	1.28	2.06	2.81	6.01	0.28	0.44
Market	2.16	3.66	0.28		1.36	3.66	0.07	

Table X. Performance statistics of the long leg of the reversal strategy for corporate bonds with monthly rebalancing. Mean is a sample average of monthly returns, in % per annum. S.D. is the standard deviation of monthly returns, in % per annum. SR is the Sharpe ratio relative to the 3 month Treasury Bill. IR is the information ratio relative to the market. The sample is Oct 2005 to Jun 2017. For portfolio construction, I apply the following filters to the sample: a) previous month outstanding amount is greater than 200 mln USD, b) previous month 12-month moving average of the realized bid-ask spread is below 100 b.p. Reversal portfolios are obtained from the double-sorting of bonds on the previous month credit rating (three terciles) and total return (five quintiles). For each of the 15 bins, the average bond return weighted by the previous month outstanding amount is computed. Long-reversal (LR) return is a simple average return across three rating terciles for the top reversal (lowest past returns) quintile. ‘LR: few funds’ is the reversal portfolio with a below-median number of fund owners. ‘LR: many funds’ is the reversal portfolio with an above-median number of fund owners. Market return is the value-weighted return of the bonds in the sample. Trading costs are assumed to be half of the 12-month average of the realized bid-ask spread (average bid-ask spread in Table I).

of the reversal strategy for bonds with more information asymmetry. Once I account for trading costs, the performance of reversal portfolios becomes considerably lower because of high portfolio turnover. However, the sub-portfolio with few fund owners still earns almost 3% per year after trading cost adjustment, which is twice more than the corporate bond market. The information ratio of the reversal portfolio with few fund owners amounts to approximately 0.5 (annualized) relative to the corporate bond market. The return on the reversal portfolio with many fund owners is considerably lower and is close the bond market after trading cost adjustment.

Figure 5 shows how reversal returns accumulate over time. Observe in Figure 5a that two-thirds of the total reversal portfolio value gains (unadjusted for trading costs) come from years 2009–2011. The difference between the value of sub-portfolios with few and many fund owners starts to accumulate since mid-2009 and is growing slowly but steadily ever since. Figure 5b plots portfolio values net of trading costs and tells a similar story except the

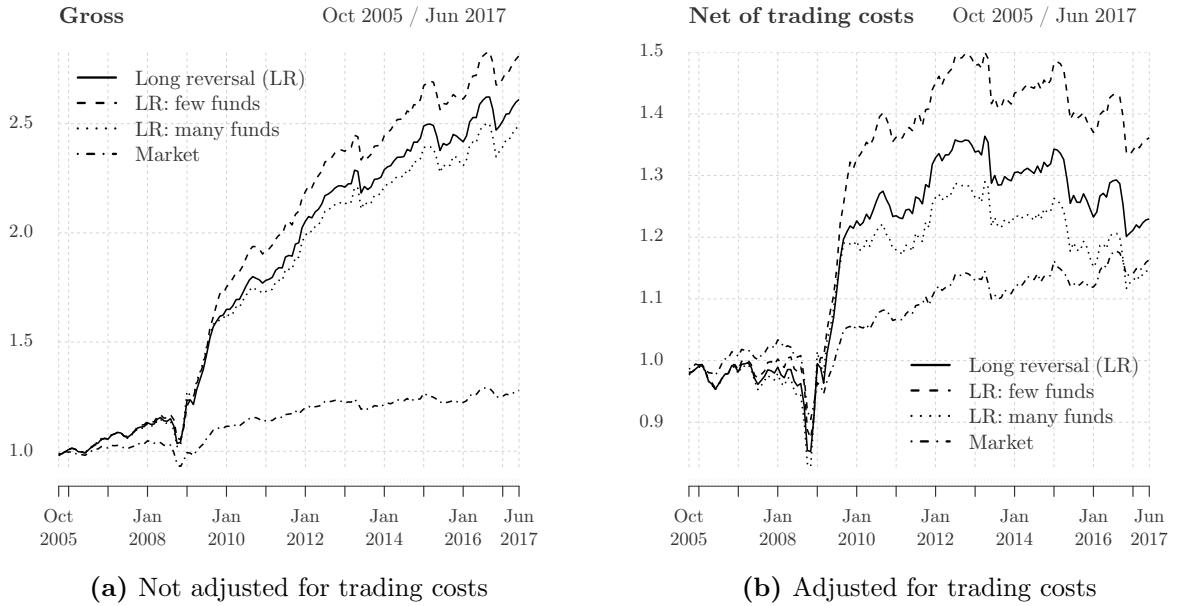


Figure 5. The value of long-reversal corporate bond portfolios with monthly rebalancing. I normalize the value of all portfolios in Sep 2005 to 1. For portfolio construction, I apply the following filters to the sample: a) previous month outstanding amount is greater than 200 mln USD, b) previous month 12-month moving average of the realized bid-ask spread is below 100 b.p. Reversal portfolios are obtained from the double-sorting of bonds on the previous month credit rating (three terciles) and total return (five quintiles). For each of the 15 bins, the average bond return weighted by the previous month outstanding amount is computed. Long-reversal (LR) return is a simple average return across three rating terciles for the top reversal (lowest past returns) quintile. ‘LR: few funds’ is the reversal portfolio with a below-median number of fund owners. ‘LR: many funds’ is the reversal portfolio with an above-median number of fund owners. Market return is the value-weighted return of the bonds in the sample. Trading costs are assumed to be half of the 12-month average of the realized bid-ask spread (average bid-ask spread in Table I).

reversal strategies here are performing worse than the market since approximately 2013. The long-reversal portfolio with few fund owners is still worth considerably more than the market portfolio by the end of the sample period.

The evidence presented in this section demonstrates that conditioning on information asymmetry considerably affects the performance of reversal strategies in practice. Reversals tend to be stronger for bonds with more information asymmetry and long-reversal portfolios with less mutual fund ownership, for instance, can outperform the corporate bond market after adjustment for trading costs. Given these findings, one can further investigate differ-

ent information asymmetry signals and potentially improve the performance of the reversal strategy on corporate bonds.

VI. The model

In this section, I present a model of competitive bond trading volume that builds on the same premises as my empirical analysis above: investors trading bonds with each other are occasionally adversely selected while dealers avoid information-driven trade flow. The model justifies equation (2), which I estimate in the empirical part of the paper, and yields predictions about the dependence of volume-return coefficients on information asymmetry that closely match empirical results I have discussed above. One can view the model of this section as the formal presentation of the intuition behind volume-return relationships I analyze in the empirical part of the paper.

The model is a modification of [Llorente et al. \(2002\)](#) which is a simplified version of [Wang \(1994\)](#) in its turn. In these models, two types of investors, informed and uninformed ones, are trading with each other for liquidity reasons and on private information. My model differs from [Llorente et al. \(2002\)](#) in two ways: I tailor the arithmetic of returns to defaultable bonds rather than to stocks as in the original model and I introduce noisy bond supply.

Changing a dividend-paying stock for a perpetual coupon-paying defaultable bond within the model requires approximations to keep the analysis tractable. In [Llorente et al. \(2002\)](#), private information is the information about dividends, which is an additive component of dollar returns. In my model, private information relates to default risk, which is not an additive term in returns calculation. To make returns linear in a default loss and simplify the learning problem for uninformed traders, I consider a log-linear approximation of defaultable bond returns as in [Hanson, Greenwood, and Liao \(2018\)](#). Given that daily bond returns in my sample are small numbers (see Table I) with 5th and 95th percentiles close to 1% in the

absolute value, the log-linear approximation of returns should not undermine the relevance of theoretical results for my empirical analysis.²³

I introduce noisy bond supply to the model to generate the additional trading volume that is not due to liquidity or informational signals the agents receive. In the model, I *assume* that supply changes that proxy for changes in dealers' bond inventory are independent of the arrival of private news. Table II suggests that such an assumption is not at odds with the data; the correlation between client-to-client and client-to-dealer daily volume measures in my sample is low. In the model, supply changes are publicly observed, unlike private liquidity signals. Under these assumptions, I can derive the dynamic volume-return relationship similar to (2) and provide additional implications for my empirical analysis compared to the baseline model of [Llorente et al. \(2002\)](#).

A. The economy

The discrete-time economy has two traded securities: a riskless bond in unlimited supply at a constant interest rate that is set to 0 for simplicity and a risky perpetual bond that pays a coupon C every period. [Hanson et al. \(2018\)](#) demonstrate that [Campbell and Shiller \(1988\)](#) decomposition applied to such a bond yields the log return r_t of the following form:

$$r_{t+1} \approx \kappa + c(1 - \theta) + \theta p_{t+1} - p_t - d_{t+1}, \quad (4)$$

where $p_t \equiv \log P_t$ is the log ex-coupon price of the bond, θ and κ are deterministic functions of the log-coupon $c \equiv \log C$, and d_{t+1} is the log default loss at time $t + 1$.²⁴

I assume that the log default loss consists of two additive components:

$$d_{t+1} = f_t + g_t.$$

f_t is publicly known at time t while g_t is a private time t information of a subset of investors.

At time $t + 1$, the value of d_{t+1} becomes publicly observed.

²³To preserve the linearity of demand with respect to state variables when working with percentage rather than dollar returns, I also have to log-linearize the wealth dynamics of the agents.

²⁴For the derivation see Appendix A.

The risky bond is traded in a competitive bond market with noisy supply s_t , which is a public knowledge. The market is populated with two classes of investors, $i = 1, 2$, with relative population weights ω and $1 - \omega$. The investors are identical within each class, and each investor's initial endowment of the risky bond is set to 0 for simplicity. Type 1 investors are informed; they observe g_t . Type 2 investors do not observe g_t but learn it from the bond price using the Bayes rule. In addition, Type 1 investors have a random exposure z_t to some non-traded asset that generates a log return of n_{t+1} in the subsequent period.²⁵ Type 2 investors do not know the exposure of type 1 investors to the non-traded risk. Overall, the information set of the informed investors at time t is $\{d, p, n, f, s, g, z\}_{0, \dots, t}$ while the information set of the uninformed investors is $\{d, p, n, f, s\}_{0, \dots, t}$.

I assume that n_t , g_t , and z_t are time-independent zero-mean normally distributed random variables with variances σ_n^2 , σ_g^2 , σ_z^2 respectively. I further assume that f_t is also time-independent and normally distributed with the mean $m_f = \kappa + c(1 - \theta)$ and the variance σ_f^2 .²⁶ All of n_t , g_t , z_t , and f_t are contemporaneously uncorrelated except for n_t and f_t that have a time-invariant negative covariance, which means that default losses are low when non-traded asset returns are high. This implies a constant positive covariance between r_t and n_t that equals σ_{rn} . Finally, the supply of the risky bond follows an AR(1) process

$$s_{t+1} = \delta s_t + \epsilon_{t+1}, \quad (5)$$

where $|\delta| < 1$ and ϵ_t is normally distributed with zero mean and variance σ_s^2 ; it is independent over time and is independent from n_t , g_t , z_t , and f_t .

The investors of both types $i = 1, 2$ maximize the next period conditional expected utility $\mathbb{E}_t \left[-e^{-W_{t+1}^{(i)}} \right]$ (derived from the next period wealth $W_{t+1}^{(i)}$ by choosing the demand $X_t^{(i)}$ for the risky bond).²⁷ To keep the model tractable I need to take the log-linear approximation

²⁵Here I follow Llorente et al. (2002) assuming for simplicity that only one type of investors has income from a non-traded asset. It is enough to generate price reversals due to liquidity trading.

²⁶The mean of f_t is chosen such that the long-term mean of the log bond price is 0 and the contributions of coupons and public news about future defaults to returns cancel one another on average.

²⁷As in Llorente et al. (2002), the risk aversion is set to 1 since it only enters the expressions for investors' demands as the multiple of the variances of all exogenous shocks. Hence, one can implement higher or lower risk aversion in the model by proportionally scaling variances of all shocks up or down.

of the wealth dynamics, which under the assumptions of the model is

$$W_{t+1}^{(1)} \approx W_t^{(1)} + X_t^{(1)} r_{t+1} + z_t(1 + n_{t+1}),$$

$$W_{t+1}^{(2)} \approx W_t^{(2)} + X_t^{(2)} r_{t+1}.$$

The model setup is different from [Llorente et al. \(2002\)](#) in two ways. First, I work with log returns approximated in (4) around $\bar{p} \equiv 0$ and linearized wealth dynamics instead of dollar returns and non-linearized wealth dynamics. Second, more importantly, I assume noisy supply (5) instead of a constant zero supply. Noisy supply allows me to decompose the trading volume in the model into two components: the first one is related to trading between informed and uninformed investors and exogenous changes in asset supply drive the second one. Empirical counterparts of these two components are respectively the volume of corporate bonds purchased by clients matched by client sales in a given period and net changes in broker-dealer inventory.

B. Model equilibrium

I solve for the rational expectations equilibrium of the model assuming a linear pricing function for the log bond price. Define the log price adjusted for the publicly known credit loss component as $\tilde{p}_t \equiv p_t + (f_t - m_f)$ and assume it is linear with respect to g_t , z_t , and s_t :

$$\tilde{p}_t = -a(g_t + bz_t + es_t). \quad (6)$$

Observe that the steady-state level of log bond price is 0 as in the linear approximation of log return (4).

Given the pricing function (6), the equation for returns (4) re-writes as:²⁸

$$r_{t+1} = -\theta(f_{t+1} - m_f) + \theta\tilde{p}_{t+1} - \tilde{p}_t - g_t. \quad (7)$$

The expression for conditional expected returns follows from (7):

$$\mathbb{E}_t^{(i)}[r_{t+1}] = -\tilde{p}_t - \mathbb{E}_t^{(i)}[g_t] - ae\theta\delta s_t.$$

²⁸In what follows, I replace an approximate equality in (4) with the exact one.

The informed investors know g_t , hence $\mathbb{E}_t^{(1)}[g_t] = g_t$. The uninformed investors observe \tilde{p}_t and s_t and estimate $\mathbb{E}_t^{(2)}[g_t|\tilde{p}_t, s_t]$. I show in Appendix A that

$$\mathbb{E}_t^{(2)}[g_t|\tilde{p}_t, s_t] = \gamma(g_t + bz_t), \quad (8)$$

where $\gamma = \frac{\sigma_g^2}{\sigma_g^2 + b^2\sigma_z^2} > 0$. One can further show that conditional return variances for two types of investors are constant over time.

With conditional expected return linear in g_t, z_t , and s_t and conditional return variance constant for both types of investors, the demand for risky bonds, $X_t^{(1)}$ and $X_t^{(2)}$, is also linear in g_t, z_t , and s_t ²⁹. The market for risky bonds clears:

$$\omega X_t^{(1)}(g_t, z_t, s_t) + (1 - \omega)X_t^{(2)}(g_t, z_t, s_t) = s_t,$$

which must hold for any values of g_t, z_t , and s_t , implying a system of three non-linear equations for yet undetermined coefficients a, b , and e . One can show that if the parameters of the model are such that the system has real-valued solutions then it must be that a, b , and e are all positive, moreover, $\omega + \gamma - \omega\gamma < a < 1$ and $b = \sigma_{rn}$. I demonstrate in Appendix A that under mild restrictions on the parameters (that boil down to σ_s^2 being not ‘too big’) the model always has real-valued solutions, of which a unique triple of $\{a^*, b^*, e^*\}$ has economically reasonable values.

C. Trading volume in the model

Consider the aggregate difference in risky bond holdings in the economy at time t

$$\omega\Delta X_t^{(1)} + (1 - \omega)\Delta X_t^{(2)} = \Delta s_t.$$

Using the equilibrium conditions one can decompose it as

$$\omega\Delta X_t^{(1)} + (1 - \omega)\Delta X_t^{(2)} = \underbrace{V_{c,t}^{(1)}(\Delta g_t, \Delta z_t) + V_{c,t}^{(2)}(\Delta g_t, \Delta z_t)}_{=0} + \underbrace{V_{s,t}^{(1)}(\Delta s_t) + V_{s,t}^{(2)}(\Delta s_t)}_{=\Delta s_t},$$

where

$$V_{c,t}^{(1)}(\Delta g_t, \Delta z_t) = V_{c,t}^{(2)}(\Delta g_t, \Delta z_t) = |\alpha(\Delta g_t + \sigma_{rn}\Delta z_t)|, \quad (9)$$

²⁹See Appendix A.

and $\alpha = \omega(a-1)/\sigma_r^2$. Here, $V_c^{(1)}$ and $V_c^{(2)}$ represent the volume of trading *between* informed and uninformed investors. This trading volume is due to changes in a private signal about credit loss Δg (information-driven trading) and the position in a non-traded asset Δz (liquidity-driven trading). $V_c^{(1)}$ and $V_c^{(2)}$ always have opposite signs but are equal in the absolute value. For the convenience of notation, I will denote this trading volume $v_{c,t} = |\alpha(\Delta g_t + \sigma_{rn}\Delta z_t)| \geq 0$. An econometrician observing bond trading records in the TRACE database can compute what the client buy volume matched by the client sell volume was at time t .³⁰ It is an empirical proxy for $v_{c,t}$.

Two other components, $V_s^{(1)}$ and $V_s^{(2)}$, represent trading due to changing bond supply. One can show that in equilibrium these two components are always of the same sign and they represent the proportion in which two types of agents absorb additional bond supply Δs . By construction, a change in bond supply is the buy volume that was not matched by the sell volume of the opposite sign. Its absolute value is equal to the absolute value of a change in aggregate dealers' inventory. The latter is an empirical counterpart of $v_{s,t} \equiv |\Delta s_t|$. What the model assumes is that $v_{c,t}$ and Δs_t are independent since the latter is uncorrelated with Δg and Δz that drive the former. Table III has demonstrated that this assumption largely holds in the data. The key takeaway of this paragraph is that I assume that an econometrician knows $v_{c,t}$ and $v_{s,t}$, and these two quantities are defined within the model as stated above.

D. Volume-return relationship and information asymmetry

Assume an econometrician observes the time-series of bond returns r_t and two types of volume, $v_{c,t}$ and $v_{s,t}$, as discussed above. Then the conditional expectation of future returns given current returns and volume can be approximated as

$$\mathbb{E}_t[r_{t+1}|r_t, v_{c,t}, v_{s,t}] \approx (\beta_1 + \beta_2 v_{c,t}^2 + \beta_3 v_{s,t}^2) \gamma_t, \quad (10)$$

³⁰All records in TRACE represent trading *between* a broker-dealer and a client and can be of two types only: a purchase by a client from a dealer or a sale to a dealer.

the derivation is presented in Appendix A. This volume-return relationship is a theoretical counterpart of equation (2) estimated in the empirical part of the paper. Unlike equation (2), equation (10) contains squared volumes. In the data, squared volumes are extremely right-skewed, hence from an econometric standpoint, it is reasonable to estimate the volume-return relationship as in (2) with volume entering the equation without a square (Llorente et al. 2002 follow the same approach). It does not change an economic interpretation of volume-return coefficients.³¹

Now, I would like to discuss how coefficients β_1 , β_2 , and β_3 change in the model as the extent of informed trading changes. In the benchmark model Llorente et al. (2002), both β_1 and β_2 are negative, but β_1 is decreasing and β_2 is increasing with the extent of information asymmetry proxied by σ_g^2 . β_1 measures the first return autocorrelation, and negative β_1 decreasing with σ_g^2 means that for two equally risky bonds returns will *revert more* for the one with more information asymmetry. β_2 measures the impact of volume on the first autocorrelation, and negative β_2 increasing with σ_g^2 means that for two equally risky bonds returns will *revert less following high-volume days* for the one with more information asymmetry. These theoretical results find empirical support in the U.S. stock market, as Llorente et al. (2002) shows.

Unlike in the benchmark model, I can not make a general statement about the signs of volume-return coefficients and their dependence on σ_g^2 ; I need to solve the model numerically first. In Figure 6, I present the relationships between information asymmetry σ_g^2 and β coefficients for the model calibrated to an average bond in TRACE. The bond has a coupon rate of 5%, high persistence of a supply shock $\delta = 0.95$, and a daily standard deviation of returns of 1%.³² The latter stays fixed in all numerical solutions; this is an additional

³¹Since an econometrician knows the sign of inventory changes, she could write an analog of equation (10) conditioning additionally on this piece of knowledge. It would change the form of the equation slightly, and the loadings on two types of volume would become incomparable. An important part of my empirical analysis consists of a direct comparison of coefficients β_2 and β_3 , and for that, I need to condition in (10) on the absolute value of inventory changes.

³²In Figure 6, I set $\delta = 0.95$ which roughly corresponds to $\text{Corr}(\Delta s_t, \Delta s_{t-1}) = -0.03$ because in the model $\text{Corr}(\Delta s_t, \Delta s_{t-1}) = -\frac{1}{2}(1 - \delta)$. In the model, δ measures the persistence of supply, which is roughly the persistence of inventory. $\delta = 0.95$ implies the half-life of broker-dealer inventory of about 13 days. Further

constraint I impose on the solutions of the model.³³ Figure 6 represents the cross-section of bonds with the same unconditional risk but different contributions of public, private, and liquidity shocks to return variance.

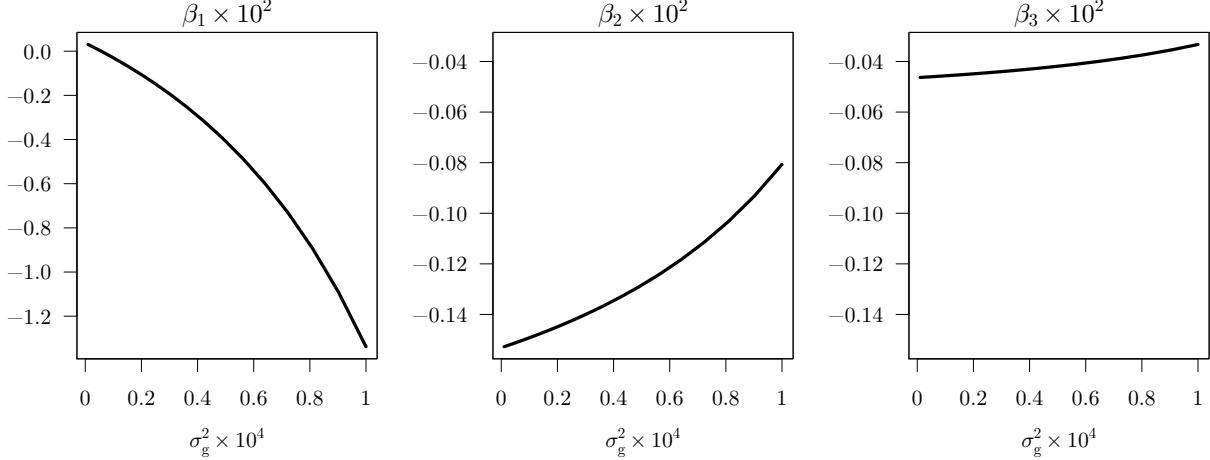


Figure 6. Dependence of β_1 , β_2 , and β_3 on information asymmetry σ_g^2 holding total return variance fixed. Each point on the curves is a numerical solution of the model. I obtain the relationships between σ_g^2 and β coefficients by varying σ_g from 0 to 1% holding an unconditional standard deviation of returns at 1%, which is a daily standard deviation of bond returns in the TRACE data. I choose the following parameters of the model to match a median bond in sample: coupon rate $C = 5\%$, the persistence of a supply shock $\delta = 0.95$. The fraction of informed investors is $\omega = 0.05$, the correlation between traded and non-traded asset returns is $\sigma_{rn} = 0.3$, the variance of the supply shock is $\sigma_s^2 = 0.1$. I first solve the model for a very small value of σ_g , 5 b.p. here. Then, I hold the equilibrium value of a fixed in all subsequent solutions for $\sigma_g > 5$ b.p.; I allow e to change. Thus, the comparative statics plotted here is a collection of solutions of the system of equations of three variables (σ_z^2 , σ_f^2 , and e): two model equilibrium equations plus an additional restriction on the total return variance.

The left and central panels in Figure 6 deliver the same message as the benchmark model. With more informed trading, returns tend to revert more, but less so following days when investors trade a lot with each other. On the left panel, which presents reversals following no-volume days, there is no reversal when σ_g is zero, and returns are due to public news that is fully priced within the same period. As σ_g increases, no-volume reversals intensify

(unreported) estimations show, in line with the results of [Dick-Nielsen and Rossi \(2018\)](#), that dealers revert deviations from their target inventory faster post-crisis.

³³[Llorente et al. \(2002\)](#) impose the same restriction on the total unconditional variance of returns.

due to a greater impact of uninformed investors' errors in estimating g_t on returns.³⁴ On the central panel, the reversal following high-volume days is the strongest when σ_g is zero because the entire trading volume between informed and uninformed investors represents, in this case, liquidity trading. Liquidity trading has price impact but does not reveal any new information about the asset payoff; hence, the price reverts the next period. As σ_g increases, it's more and more likely that some part of the between-investors trading volume comes from Δg and conveys the information about future returns; hence the reversal tends to decrease (β_2 tends to increase). The right panel in Figure 6 shows that β_3 that measures an additional component of reversals following days when inventory changes a lot is relatively insensitive to σ_g . It does not look surprising given that Δs in the model is uncorrelated with other motives for trading. One would expect β_3 to be flat with respect to σ_g in such case; a slightly upward sloping line on the right panel of Figure 6 is due to equilibrium e (price impact of inventory-changing trades) changing with σ_g .

The shape of the lines in Figure 6 matches closely the shape of their empirical counterparts presented in Figure 3. In the model, as it is in the data, β_1 decreases, and β_2 increases with information asymmetry, while β_3 is insensitive to information asymmetry. It gives additional support for the premises of the model: client-to-client trading volume may be due to private information, but client-to-dealer trading volume is likely driven by liquidity needs only.

As in Llorente et al. (2002), the limitation of my extended model is that β_2 stays negative for all reasonable model calibrations and does not turn positive (same applies to β_3 which is not the part of the benchmark model). In reality, as Section III has shown, β_2 is positive for most corporate bonds. It does not undermine the main idea suggested by the model and tested in the empirical part of the paper. As the extent of informed trading increases,

³⁴Here is the intuition for this result. With no volume, time t returns are not driven by liquidity shocks since Δz_t and Δg_t must be zero. Assume $z_{t-1} > 0$ and informed investors are net sellers of bonds. From (7) and (8) one finds that r_t is negative when $\frac{a}{\gamma} \mathbb{E}_{t-1}^{(2)}[g_{t-1}] < g_{t-1}$ other things being equal, i.e., when actual losses in default are higher than previously expected by uninformed investors. But that means that in $t-1$ informed investors' demand for bonds was lower than required by their hedging needs; so it is in t since the volume is zero. Hence, time t price is low and time $t+1$ expected return is high. Higher information asymmetry amplifies this effect.

returns following high-volume days are less likely to revert, especially when dealers are not trading from their inventory capacity.

VII. Conclusion

In this paper, I estimate a dynamic volume-return relationship for individual bonds and explore the determinants of estimated volume-return coefficients in a cross-section of bonds. A particular focus of my analysis is on the impact of information asymmetry on volume-return coefficients.

The hypotheses that I test arise from a stylized theoretical model of competitive bond trading with asymmetric information and non-traded risks. In the model, trading between investors is due to liquidity needs (hedging of the non-traded risk) or private information. Also, investors in the model absorb random bond supply shocks; their empirical counterpart is the change in aggregate bond inventory. The model suggests that bonds with high information asymmetry have stronger price reversals than bonds with low information asymmetry, but less so following high-volume days when dealers' inventory *does not change*, and investors are essentially trading with each other. Conversely, following days with substantial changes in dealers' inventory, the difference in reversals between high- low-asymmetry bonds remains. In the model, this result emerges because changes in inventory (supply shocks) are assumed independent from the arrival of private news.

I find strong empirical support for model predictions in the data. Bonds with high information asymmetry exhibit stronger price reversals than low-asymmetry bonds, but less so following days when trading volumes are high, but dealers' inventory does not change at the end of the day (clients purchases equal client sales). High-asymmetry bonds in my analysis are the bonds that are owned by few mutual funds and intermediated by few dealers, have smaller outstanding amounts and issued by smaller firms with no actively traded CDS contract on the issuer and high stock return volatility.

In particular, I find that a typical bond with high information asymmetry has the first autocorrelation of returns close to -0.4 following average-volume trading days. Following two standard deviations above-average volume day when dealers' inventory does not change, the first autocorrelation reduces to -0.18. A similar bond with *the same* average realized bid-ask spread, return volatility, credit spread, and volume autocorrelation, but low information asymmetry has the first return autocorrelation of -0.2, which increases only by 0.05 to -0.15 following high-volume inventory-neutral days.

If one considers, instead, the reversals following days when trading volume is high, but it is due to substantial changes in dealers' inventory, then the difference in reversals between bonds with high and low information asymmetry remains at the average-volume day level. These results are consistent with the assumption that trading volume in high-asymmetry bonds is more likely to come from investors who possess private information. Since dealers typically know their clients well and might be able to detect informed investors, they let other investors provide liquidity for such trades. Overall, my results suggest that there might be informed trading in corporate bonds, but when it happens, dealers are not providing liquidity and are not adversely selected.

My findings have implications for the design of investment strategies exploiting corporate bond reversals. In particular, I show that long-reversal portfolios of high-asymmetry bonds outperform long-reversal portfolios of low-asymmetry bonds both before and after adjustment for trading costs. Hence, illiquidity does not fully explain reversal returns. Moreover, reversal portfolios of high-asymmetry bonds outperform the corporate bond market after trading cost adjustment. An investor considering an implementation of a bond reversal strategy might profit from additionally sorting bonds on information asymmetry proxies.

My results also relate to a recent policy debate about corporate bond market transparency. I find that bonds with less transparent valuations tend to have stronger price reversals when trading is purely liquidity-driven, and fundamental values of the bonds likely remain unchanged. Stronger liquidity-driven reversal is just another name for non-fundamental

price volatility that is often regarded as an undesirable feature of a well-functioning financial market. From this standpoint, a proposed reduction in corporate bond market transparency (TRACE delayed trade dissemination pilot project) might not be optimal.

REFERENCES

Adrian, Tobias, Nina Boyarchenko, and Or Shachar, 2017, Dealer balance sheets and bond liquidity provision, *Journal of Monetary Economics* 89, 92 – 109.

Asquith, Paul, Andrea S. Au, Thomas Covert, and Parag A. Pathak, 2013, The market for borrowing corporate bonds, *Journal of Financial Economics* 107, 155 – 182.

Bai, Jennie, Turan G. Bali, and Quan Wen, 2019, Common risk factors in the cross-section of corporate bond returns, *Journal of Financial Economics* 131, 619 – 642.

Bali, Turan G., Avanidhar Subrahmanyam, and Quan Wen, 2018, Long-term contrarian strategies in the corporate bond market, Working paper, available at
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2978861.

Bao, Jack, Maureen O’Hara, and Xing (Alex) Zhou, 2018, The Volcker Rule and corporate bond market making in times of stress, *Journal of Financial Economics* 130, 95 – 113.

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

Barbon, Andrea, Marco Di Maggio, Francesco A. Franzoni, and Augustin Landier, 2018, Brokers and order flow leakage: Evidence from fire sales, Swiss Finance Institute Working Paper no. 17-61, available at
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2991617.

Berndt, Antje, and Yichao Zhu, 2018, Dealer inventory, short interest and price efficiency in the corporate bond market, Working paper, available at
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3189980.

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

BlackRock, 2018, The next generation bond market, SEC Fixed Income Market Structure Advisory Committee (FIMSAC) meeting memo, Available at

<https://www.sec.gov/spotlight/fixed-income-advisory-committee/blackrock-next-generation-bond-market-fimsa-011118.pdf>.

Campbell, John Y., Sanford J. Grossman, and Jiang Wang, 1993, Trading volume and serial correlation in stock returns, *Quarterly Journal of Economics* 108, 905–939.

Campbell, John Y., and Robert J. Shiller, 1988, Stock prices, earnings, and expected dividends, *The Journal of Finance* 43, 661–676.

Choi, Jaewon, and Yesol Huh, 2018, Customer liquidity provision: Implications for corporate bond transaction costs, Working paper, available at

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2848344.

Chordia, Tarun, Amit Goyal, Yoshio Nozawa, Avanidhar Subrahmanyam, and Qing Tong, 2017, Are capital market anomalies common to equity and corporate bond markets? An empirical investigation, *Journal of Financial and Quantitative Analysis* 52, 1301–1342.

Citi, 2018, Developments in credit market liquidity, SEC Fixed Income Market Structure Advisory Committee (FIMSAC) meeting memo, Available at

<https://www.sec.gov/spotlight/fixed-income-advisory-committee/citi-developments-in-credit-market-liquidity-fimsa-011118.pdf>.

Dick-Nielsen, Jens, 2014, How to clean Enhanced TRACE data, Working paper, available at

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2337908.

Dick-Nielsen, Jens, and Marco Rossi, 2018, The cost of immediacy for corporate bonds, *The Review of Financial Studies* 32, 1–41.

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

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

Feldhütter, Peter, and Thomas K. Poulsen, 2018, What determines bid-ask spreads in over-the-counter markets?, Working paper, available at
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286557.

FINRA, 2019, Proposed pilot program to study recommended changes to corporate bond block trade dissemination, Regulatory notice no. 19-12, Available at
<https://www.finra.org/industry/notices/19-12>.

Friewald, Nils, and Florian Nagler, 2019, Over-the-counter market frictions and yield spread changes, *The Journal of Finance* .

Gilchrist, Simon, and Egon Zakrajšek, 2012, Credit spreads and business cycle fluctuations, *American Economic Review* 102, 1692–1720.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71 – 100.

Goldstein, Michael A., and Edith S. Hotchkiss, 2019, Providing liquidity in an illiquid market: Dealer behavior in us corporate bonds, *Journal of Financial Economics* .

Han, Song, and Xing Zhou, 2014, Informed bond trading, corporate yield spreads, and corporate default prediction, *Management Science* 60, 675–694.

Hanson, Samuel G, Robin Greenwood, and Gordon Y Liao, 2018, Asset price dynamics in partially segmented markets, *The Review of Financial Studies* 31, 3307–3343.

Hendershott, Terrence, Roman Kozhan, and Vikas Raman, 2019, Short selling and price discovery in corporate bonds, *Journal of Financial and Quantitative Analysis* 139.

Ivashchenko, Alexey, and Artem Neklyudov, 2018, (In)frequently traded corporate bonds, Working paper, available at
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3306124.

Llorente, Guillermo, Roni Michaely, Gideon Saar, and Jiang Wang, 2002, Dynamic volume-return relation of individual stocks, *The Review of Financial Studies* 15, 1005–1047.

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

Schestag, Raphael, Philipp Schuster, and Marliese Uhrig-Homburg, 2016, Measuring liquidity in bond markets, *The Review of Financial Studies* 29, 1170–1219.

Wang, Jiang, 1994, A model of competitive stock trading volume, *Journal of Political Economy* 102, 127–68.

Zawadowski, Adam, and Martin Oehmke, 2016, The anatomy of the CDS Market, *The Review of Financial Studies* 30, 80–119.

Appendix A. Aspects of the model

Log-linear approximation of returns

Consider a homogeneous portfolio of perpetual defaultable bonds with invoice price P_t and coupon rate C . Its next period return R_{t+1} is:

$$1 + R_{t+1} = \frac{(1 - D_{t+1})(P_{t+1} + C)}{P_t},$$

where $D_{t+1} = h_{t+1}L_{t+1}$, and h_{t+1} represents a default rate and $L_{t+1} \in [0, 1]$ represents loss given default for bonds in the portfolio at time $t + 1$.³⁵ Define $r_t \equiv \log(1 + R_t)$, $p_t \equiv \log(P_t)$, $c \equiv \log(C)$, and $-d_t \equiv \log(1 - D_t)$. Then

$$\begin{aligned} r_{t+1} &= -d_{t+1} - p_t + \log(P_{t+1} + C) \\ &= -d_{t+1} - p_t + p_{t+1} + \log\left(1 + \frac{C}{P_{t+1}}\right) \\ &= -d_{t+1} - p_t + p_{t+1} + \log\left(1 + e^{c-p_{t+1}}\right) \end{aligned}$$

Notice that the first-order Taylor expansion of $\log(1 + e^{c-x})$ around $c - \bar{x}$ yields:

$$\log(1 + e^{c-x}) \approx \log(1 + e^{c-\bar{x}}) + \frac{e^{c-\bar{x}}}{1 + e^{c-\bar{x}}} ((c - x) - (c - \bar{x})).$$

Then the expression for returns becomes:

$$\begin{aligned} r_{t+1} &= -d_{t+1} - p_t + p_{t+1} + \log\left(1 + e^{c-\bar{p}_{t+1}}\right) + \underbrace{\frac{e^{c-\bar{p}}}{1 + e^{c-\bar{p}}}(c - p_{t+1}) - \frac{e^{c-\bar{p}}}{1 + e^{c-\bar{p}}}(c - \bar{p})}_{\text{Call } \theta = \frac{1}{1 + e^{c-\bar{p}}} \Rightarrow \frac{e^{c-\bar{p}}}{1 + e^{c-\bar{p}}} = 1 - \theta} \\ &= -d_{t+1} - p_t + p_{t+1} - \log\theta + (1 - \theta)(c - p_{t+1}) - (1 - \theta)(c - \bar{p}) \\ &= \theta p_{t+1} - p_t - d_{t+1} + (1 - \theta)c + \underbrace{(-\log\theta - (1 - \theta)\log(\theta^{-1} - 1))}_{\equiv \kappa} \end{aligned}$$

which is equation (4). I set $\bar{p} = 0$ (the steady-state bond price is par), then $\theta = \frac{1}{1+C}$.

³⁵With probability $1 - h_{t+1}$ the bond pays $P_{t+1} + C$ and with probability h_{t+1} it pays $(1 - L_{t+1})(P_{t+1} + C)$.

Learning by uninformed investors

The uninformed investor is a Bayesian agent learning about g_t and z_t at time t by observing \tilde{p}_t and s_t . Recall that

$$\tilde{p}_t = -a(g_t + bz_t + es_t).$$

Hence, the agent knows $g_t + bz_t$ and an estimate of g_t immediately gives an estimate of z_t .

The conditional distribution of \tilde{p}_t given g_t and s_t is

$$\tilde{p}_t | g_t, s_t \sim N(-a(g_t + es_t), a^2 b^2 \sigma_z^2).$$

The unconditional distribution of g_t is $N(0, \sigma_g^2)$. Bayes theorem implies that $g_t | \tilde{p}_t, s_t$ is also Normal with a PDF $f_{g|\tilde{p},s}$:

$$f_{g|\tilde{p},s} \propto \exp \left(\underbrace{-\frac{(\tilde{p}_t + a(g_t + es_t))^2}{2a^2 b^2 \sigma_z^2} - \frac{g_t^2}{2\sigma_g^2}}_{\equiv -\frac{1}{2}K} \right).$$

Expanding the square and collecting terms, one gets:

$$K = \frac{g_t^2 - 2g_t \left[-\frac{a\sigma_g^2 \tilde{p}_t + a^2 \sigma_g^2 es_t}{a^2 (\sigma_g^2 + b^2 \sigma_z^2)} \right] + \Lambda(\tilde{p}_t, s_t)}{\frac{b^2 \sigma_z^2 \sigma_g^2}{\sigma_g^2 + b^2 \sigma_z^2}},$$

where $\Lambda(\tilde{p}_t, s_t)$ does not depend on g_t . Plug in the expression for the pricing function

$\tilde{p}_t = -a(g_t + bz_t + es_t)$ to get:

$$\begin{aligned} \mathbb{E}_t^{(2)}[g_t | \tilde{p}_t, s_t] &= \underbrace{\frac{\sigma_g^2}{\sigma_g^2 + b^2 \sigma_z^2}}_{\equiv \gamma} (g_t + bz_t), \\ \mathbb{V}_t^{(2)}[g_t | \tilde{p}_t, s_t] &= (1 - \gamma) \sigma_g^2. \end{aligned}$$

Optimal demands

The informed investor is solving the following problem:

$$\max_{X_t^{(1)}} \mathbb{E}_t \left[e^{-\left(W_t^{(1)} + X_t^{(1)} r_{t+1} + Z_t (1 + n_{t+1}) \right)} \right]$$

where the distributions of r_{t+1} and n_{t+1} given the informed investor's information set at time t are both Normal with means $\mathbb{E}_t^{(1)}[r_{t+1}]$ and 0, and variances $\mathbb{V}_t^{(1)}[r_{t+1}]$ and σ_n^2 correspondingly. The covariance between r_{t+1} and n_{t+1} is time-invariant and equals σ_{rn} by assumption.

The solution of the informed investor's optimization problem is

$$X_t^{(1)} = \frac{\mathbb{E}_t^{(1)}[r_{t+1}] - \sigma_{rn}Z_t}{\mathbb{V}_t^{(1)}[r_{t+1}]}.$$

The optimization problem for the uninformed investor (who does not own the non-traded asset by assumption) is the same up to Z_t component in the wealth dynamic and yields

$$X_t^{(2)} = \frac{\mathbb{E}_t^{(2)}[r_{t+1}]}{\mathbb{V}_t^{(2)}[r_{t+1}]}.$$

Conditional variances $\mathbb{V}_t^{(1)}[r_{t+1}]$ and $\mathbb{V}_t^{(2)}[r_{t+1}]$ are constant:

$$\begin{aligned}\mathbb{V}_t^{(1)}[r_{t+1}] &= \theta^2(\sigma_f^2 + \sigma_{\bar{p}}^2), \\ \mathbb{V}_t^{(2)}[r_{t+1}] &= \theta^2(\sigma_f^2 + \sigma_{\bar{p}}^2) + (1 - \gamma)\sigma_g^2,\end{aligned}$$

Now, call $\sigma_r^2 \equiv \theta^2(\sigma_f^2 + \sigma_{\bar{p}}^2)$ and plug in the expressions for conditional expected returns and variances into the expressions for optimal demand to get:

$$\begin{aligned}X_t^{(1)} &= \frac{a-1}{\sigma_r^2}g_t + \frac{b(a-1)}{\sigma_r^2}z_t + \frac{ae(1-\theta\delta)}{\sigma_r^2}s_t, \\ X_t^{(2)} &= \frac{a-\gamma}{\sigma_r^2 + (1-\gamma)\sigma_g^2}g_t + \frac{b(a-\gamma)}{\sigma_r^2 + (1-\gamma)\sigma_g^2}z_t + \frac{ae(1-\theta\delta)}{\sigma_r^2 + (1-\gamma)\sigma_g^2}s_t.\end{aligned}$$

Existence of the equilibrium

The equilibrium conditions imply the following system of three non-linear equations in a , b , and e :

$$\begin{aligned}\frac{\omega(a-1)}{\sigma_r^2} + \frac{(1-\omega)(a-\gamma)}{\sigma_r^2 + (1-\gamma)\sigma_g^2} &= 0, \\ \frac{\omega(ab-\sigma_{rn})}{\sigma_r^2} + \frac{(1-\omega)(a-\gamma)b}{\sigma_r^2 + (1-\gamma)\sigma_g^2} &= 0, \\ \frac{\omega ae(1-\theta\delta)}{\sigma_r^2} + \frac{(1-\omega)ae(1-\theta\delta)}{\sigma_r^2 + (1-\gamma)\sigma_g^2} &= 1.\end{aligned}$$

The second equation immediately implies that $b = \sigma_{rn}$ is the only possible solution for b .

The system of two remaining equations for a and e can be re-written as

$$0 = \phi_1(a, e) \equiv (a - \bar{a})(\sigma_r^2 + \omega(1 - \gamma)\sigma_g^2) - (1 - \bar{a})\omega(1 - \gamma)\sigma_g^2,$$

$$0 = \phi_2(a, e) \equiv ae(1 - \theta\delta)\omega(1 - \gamma) - \sigma_r^2(a - \gamma),$$

where $\bar{a} = \omega + \gamma - \omega\gamma > \gamma > 0$. Observe from the first equation that $\phi_1(\bar{a}, e) < 0$ and $\phi_1(1, e) > 0$. Hence, if the solution a^* exists, it must be that $a^* \in (\bar{a}, 1)$. Then, take the derivative of the first equation with respect to a treating e as a function of a :

$$\frac{d}{da} [\phi_1(a, e(a))] = \sigma_r^2 + \omega(1 - \gamma)\sigma_g^2 + (a - \bar{a})(\sigma_g^2 + b^2\sigma_z^2 + \sigma_s^2e^2 + \sigma_s^2ae\frac{d}{da}[e(a)]),$$

which is positive for $a \in (\bar{a}, 1)$ if $e^*(a)$ that solves the second equation $0 = \phi_2(a, e)$ grows in a . In this case we would have a unique positive solution $a^* \in (\bar{a}, 1)$. Now, I am going to establish the conditions under which this is indeed the case.

The second equation can be re-written as a quadratic equation with respect to e :

$$0 = \phi_2(a, e) = (a^2(a - \gamma)\theta^2\sigma_s^2) e^2 - (a(1 - \theta\delta)\omega(1 - \gamma)) e + (a - \gamma)\theta^2(\sigma_f^2 + a^2(\sigma_g^2 + b^2\sigma_z^2)).$$

Since $a^* > \bar{a} > \gamma$, it must be that $\phi_2(a, 0) > 0$, and if the solution e^* exists it must be that $e^* > 0$. Two candidate solutions of the quadratic equation can be written as:

$$\begin{aligned} e^*(a) &= v(a) \pm v(a)k(a) \text{ where} \\ v(a) &\equiv \underbrace{\frac{(1 - \theta\delta)(1 - \gamma)\omega}{2\theta^2\sigma_s^2}}_{\equiv 1/B} \frac{1}{a(a - \gamma)}, \\ k(a) &\equiv \sqrt{1 - B^2\psi(a)}, \\ \psi(a) &\equiv (a - \gamma)^2 \left(\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2}a^2 \right) \end{aligned}$$

and for $a \in (\bar{a}, 1)$ $v > 0, v' < 0, 0 < k < 1, k' < 0, \psi > 0, \psi' > 0$. For the solutions to exist it must be that $\psi < B^{-2}$ for $a \in (\bar{a}, 1)$. Observe that

$$\begin{aligned} \psi &= (a - \gamma)^2 \left(\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_f^2}{\sigma_s^2}a^2 \right) \left(1 - \gamma \right)^2 \left(\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_f^2}{\sigma_s^2}a^2 \right) \text{ (and)} \\ B^{-2} &= \frac{(1 - \theta\delta)^2(1 - \gamma)^2\omega^2}{4\theta^4\sigma_s^4}. \end{aligned}$$

So, it is suffice to impose the following restriction on model parameters:

$$\frac{(1-\theta\delta)^2\omega^2}{4\theta^4} \frac{1}{\sigma_s^2(\sigma_f^2 + \sigma_g^2 + b^2\sigma_z^2)} > 1,$$

to guarantee that the discriminant is non-negative and the quadratic equation for e has solutions. The condition is easy to obey since the shocks in the left-hand side denominator are small numbers. From now on I assume that the condition is satisfied.

Of the two roots of the quadratic equation for e , I am going to focus on the smaller one, $e^*(a) = v(a) - v(a)k(a)$. First, it is the root that guarantees that $e^*(a)$ grows with a when $a \in (\bar{a}, 1)$ as I am about to prove. Second, for reasonable parameters values $v(a)$ is a fairly large number (in a numerical example in Section VI it is around 60) and a positive root $v(a) + v(a)k(a)$ does not make much economic sense.

The smaller root $e^*(a) = v(a) - v(a)k(a)$ grows with $a \in (\bar{a}, 1)$ if $\frac{d}{da}[e^*(a)] > 0$, i.e.:

$$v' - v'k - vk' > 0 \Leftrightarrow$$

$$v'(1 - k) > vk' \Leftrightarrow$$

$$\begin{aligned} \frac{v'}{v} &> \frac{k'}{1 - k} \Leftrightarrow \\ \frac{v'}{v} &> \frac{k'(1 + k)}{1 - k^2} \Leftrightarrow \\ \frac{v'}{v} &> \frac{-\frac{1}{2k}B^2\psi'(1 + k)}{B^2\psi} \Leftrightarrow \\ \frac{v'}{v} &> -\frac{1}{2} \frac{\psi'(1 + \frac{1}{k})}{\psi} \Leftrightarrow \\ -\frac{2a - \gamma}{a(a - \gamma)} &> -\frac{1}{2} \frac{\psi'(1 + \frac{1}{k})}{\psi} \Leftrightarrow \\ -\frac{2a - \gamma}{a(a - \gamma)} &> -\frac{\frac{\sigma_f^2}{\sigma_s^2}(a - \gamma) + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2}a(a - \gamma)(2a - \gamma)}{(a - \gamma)^2 \left(\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2}a^2 \right)} \left(1 + \frac{1}{k} \right) \Leftrightarrow \\ 2 - \frac{\gamma}{a} &< \frac{\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2}a(2a - \gamma)}{\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2}a^2} \left(1 + \frac{1}{k} \right) \end{aligned}$$

(and observe that

$$2 - \frac{\gamma}{a} < 2 < 1 + \frac{1}{k} < \frac{\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2}a(2a - \gamma)}{\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2}a^2} \left(\left(1 + \frac{1}{k} \right), \right.$$

which is indeed true.

To sum up, under the condition

$$\frac{(1 - \theta\delta)^2\omega^2}{4\theta^4} \frac{1}{\sigma_s^2 (\sigma_f^2 + \sigma_g^2 + b^2\sigma_z^2)} > 1$$

the equation $0 = \phi_2(a, e)$ always has a root $e^*(a) > 0$ that grows with $a \in (\bar{a}, 1)$, and it leads to the unique solution $a^* \in (\bar{a}, 1)$ of $0 = \phi_1(a, e^*(a))$.

Derivation of the volume-return relationship

Plug in the expression for the pricing function $\tilde{p}_t = -a(g_t + bz_t + es_t)$ into (7) to get

$$r_t = -\theta(f_t - m_f) - a\theta g_t - a\theta bz_t - a\theta es_t + (a - 1)g_{t-1} + abz_{t-1} + aes_{t-1}.$$

Assume an econometrician also observes $v_{c,t} = |\alpha(\Delta g_t + \sigma_{rn}\Delta z_t)|$ and $v_{s,t} = s_t - s_{t-1}$. Now, the goal is to compute $\mathbb{E}_t [r_{t+1}|r_t, v_{c,t}, v_{s,t}]$.

Call, for the sake of convenience of notations, $x \equiv r_{t+1}$, $y \equiv r_t$, $v \equiv \alpha(\Delta g_t + \sigma_{rn}\Delta z_t)$, and $u \equiv v_{s,t}$. The unconditional distribution of (x, y, v, u) is Gaussian:

$$(x, y, v, u)' \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ \vdots \\ \vdots \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \right)$$

where $\Sigma_{11} = \sigma_{xx}$, $\Sigma_{12} \equiv [\sigma_{xy} \ \sigma_{xv} \ \sigma_{xu}]$ and

$$\Sigma_{22} \equiv \begin{pmatrix} \sigma_{yy} & \sigma_{yv} & \sigma_{yu} \\ \sigma_{yv} & \sigma_{vv} & 0 \\ \sigma_{yu} & 0 & \sigma_{uu} \end{pmatrix}.$$

The projection theorem for multivariate Normal distributions implies:

$$\mathbb{E}[x|y, v, u] = \beta_{xy}y + \beta_{xv}v + \beta_{xu}u,$$

where $(\beta_{xy} \ \beta_{xv} \ \beta_{xu}) = \Sigma_{12}\Sigma_{22}^{-1}$.

Now consider $\mathbb{E}[x|y, |v|, u]$. First, apply the law of iterated expectations:

$$\begin{aligned}\mathbb{E}[x|y, |v|, u] &= \mathbb{E}[\mathbb{E}[x|y, v, u] |y, |v|, u] \\ &= \mathbb{E}[\beta_{xy}y + \beta_{xv}v + \beta_{xu}u |y, |v|, u] \\ &= \beta_{xy}y + \beta_{xv}\mathbb{E}[v|y, |v|, u] + \beta_{xu}u.\end{aligned}$$

Notice that $\mathbb{E}[v|y, |v|, u] = \mathbb{E}[v|y, |v|]$ since $\sigma_{vu} = 0$. Now, use the fact that for any random variable Q with a PDF $f_Q(q)$:

$$\mathbb{E}[Q|q] = |q| \frac{f_Q(|q|) - f_Q(-|q|)}{f_Q(|q|) + f_Q(-|q|)}.$$

In this case, it implies:

$$\mathbb{E}[v|y, |v|] = |v| \frac{f_{v|y}(|v|) - f_{v|y}(-|v|)}{f_{v|y}(|v|) + f_{v|y}(-|v|)},$$

where

$$v|y \sim \mathcal{N}\left(\frac{\sigma_{yv}}{\sigma_{yy}}y, \sigma_{vv} - \frac{\sigma_{yv}^2}{\sigma_{yy}}\right)$$

After straightforward algebra, one finds that

$$\mathbb{E}[v|y, |v|] = |v| \frac{e^{\rho|v|y} - e^{-\rho|v|y}}{e^{\rho|v|y} + e^{-\rho|v|y}} \approx \rho_{yv}|v|^2y$$

for small values of v , where $\rho_{yv} = \frac{\sigma_{yv}}{\sigma_{vv}\sigma_{yy} - \sigma_{yv}^2}$.

Assembling altogether:

$$\mathbb{E}[x|y, |v|, u] \approx (\beta_{xy} + \rho\beta_{xv}|v|^2) \underbrace{y}_{\text{if } v \neq 0} + \beta_{xu}u.$$

Since v and u are assumed independent, an additional conditioning on $|u|$ in the expectation sign is straightforward:

$$\mathbb{E}[x|y, |v|, |u|] \approx (\beta_{xy} + \rho_{yv}\beta_{xv}|v|^2 + \rho_{yu}\beta_{xu}|u|^2) \underbrace{y}_{\text{if } v \neq 0},$$

which is the analogue of (10). Above, $\rho_{yu} = \frac{\sigma_{yu}}{\sigma_{uu}\sigma_{yy} - \sigma_{yu}^2}$. To compute the coefficients in this relationship given model parameters one needs to compute the covariance matrix Σ . Direct

calculations yield:

$$\begin{aligned}
\sigma_{xx} &= \theta^2 \sigma_f^2 + ((a\theta)^2 + (a-1)^2) \sigma_g^2 + (ab)^2(\theta^2 + 1) \sigma_z^2 + \frac{(ae)^2(\theta^2 + 1 - 2\theta\delta)}{1 - \delta^2} \sigma_s^2; \\
\sigma_{xy} &= (1-a)a\theta\sigma_g^2 - (ab)^2\theta\sigma_z^2 + \frac{(ae)^2(\theta\delta(1-\delta) + \delta - \theta)}{1 - \delta^2} \sigma_s^2; \\
\sigma_{xv} &= \alpha(a(\sigma_g^2 + b^2\sigma_z^2) - \sigma_g^2); \\
\sigma_{xu} &= \frac{ae(1 - \theta\delta)}{1 + \delta} \sigma_s^2; \\
\sigma_{yy} &= \sigma_{xx}; \\
\sigma_{yv} &= \alpha(1 - a(1 + \theta))\sigma_g^2 - \alpha ab^2(1 + \theta)\sigma_z^2; \\
\sigma_{yu} &= -\frac{ae(1 + \theta)}{1 + \delta} \sigma_s^2; \\
\sigma_{vv} &= 2\alpha^2(\sigma_g^2 + b^2\sigma_z^2); \\
\sigma_{uu} &= \frac{2}{1 + \delta} \sigma_s^2.
\end{aligned}$$

Appendix B. Data and sample

Sample selection

I apply some filters to the TRACE database *after* cleaning it as in [Dick-Nielsen \(2014\)](#).

Here are the criteria I use to select the bonds in the sample:

- The bond is nominated in USD;
- It is a fixed coupon (including zero-coupon), non-asset backed, non-convertible, non-enhanced bond;
- Not privately issued and not issued under Rule 144A;
- Of one of the following types according to the Mergent FISD classification: CMTN (US Corporate MTN), CDEB (US Corporate Debentures), CMTZ (US Corporate MTN Zero), CZ (US Corporate Zero), USBN (US Corporate Bank Note), PS (Preferred Security), UCID (US Corporate Insured Debenture);
- The interest is paid 1, 2, 4, or 12 times a year, or the bond is zero-coupon;
- The quoting convention is 30/360.

Four additional criteria must be jointly satisfied to keep a trade record in the sample:

- The trade is executed between Jan 1, 2010, and Jun 30, 2017;
- Executed at eligible times (time stamps of the trades are between 00:00:00 and 23:59:59; there is a small number of trades in TRACE with misreported times that do not fall into this range, I remove them from the sample);
- Executed on NYSE business days;
- Executed on or after the dated date of the bond (the date when the interest starts to accrue).

Agency transactions with commissions are retained in the sample.

Winsorization

To ensure that my results are not driven by extreme observations, I winsorize some variables. In particular, in the original bond-day panel (before active periods are determined) I winsorize:

- C-to-C trading volume at 99%;
- C-to-D trading volume at 1% and 99%;
- Credit spread at 99.9%;
- Bid-ask spread at 99.9%;
- Total daily returns at 0.1% and 99.9%.

SEC N-Q forms and holdings data

Adapted from [Ivashchenko and Neklyudov \(2018\)](#).

SEC N-Q forms submitted by mutual funds are available online through the SEC EDGAR system. I recover the number of mutual fund owners from machine-read N-Q forms. Mutual funds have a lot of discretion in how they fill in their N-Q forms, which makes it challenging to process the data. Below I describe the main steps I take.

Funds do not report bond CUSIP numbers in N-Q forms. I identify bond holdings in N-Q forms by issuer name, maturity, and coupon rate. I attempt to find N-Q records matching the CUSIPs in my sample. Several possibilities arise. If there is no match, I assign a value of zero to the number of mutual funds that hold the bond. If there is a match, it may or may not be unique. Even if the match is unique (which is the dominant case), it may refer to a not-in-sample bond with the same coupon rate and maturity. To ensure that I identify a correct bond I compute the cosine text similarity measure between an issuer name from the FISD database and an issuer name I recover from N-Q forms.³⁶ Table B1 provides some examples. Table B1a shows a record with a uniquely identified bond, while Table B1b shows

³⁶I experimented with different similarity measures and did not observe much difference in results.

a record with double matching: one bond is the true bond I am looking for, another bond is mortgage-backed security with the same coupon rate and maturity. Regardless of whether the match is unique or not, I keep a record in the sample only if the cosine similarity measure is above 0.45.

cusip_id	issuer	maturity	rate	report	CIK	what	similarity
22541LAL7	credit suisse first boston (usa) inc	2009-01-15	3.88	2005-01-31	0000933996	credit suisse fb usa inc	0.67

(a) Unique maturity and coupon rate pair

cusip_id	issuer	maturity	rate	report	CIK	what	similarity
36158FAA8	ge global ins hldg corp	2026-02-15	7.00	2005-01-31	0000933996	ge global insurance holding	0.56
36158FAA8	ge global ins hldg corp	2026-02-15	7.00	2005-01-31	0000933996	fhlmc pool	0.17

(b) Non-unique maturity and coupon rate pair

Table B1. Examples of records with a unique and non-unique combination of maturity and coupon rate. First four columns (CUSIP number, issuer name, maturity, and coupon rate) are the data from Mergent FISD. The next three columns (report date, investment fund identifier CIK, and ‘what’) are the data from a machine-read N-Q filing matched to the FISD data by maturity and coupon rate. ‘Similarity’ is a cosine similarity between ‘issuer’ and ‘what’ fields.

Funds report N-Q forms twice every fiscal year, which is fund-specific. I use the ‘last observation carried forward’ approach to fill business days between two reporting dates for every fund, i.e., I assume that a fund holds all recently reported bonds every business day between the most recent and the previous N-Q report.³⁷ Once I have fund-bond-day holding indicators, I compute the number of funds that hold the bond in a given period and use this variable as the ‘number of mutual fund owners’ in my analysis.

Actively traded CDS contracts

DTCC publishes a list of 1000 most actively traded single-name CDS contracts quarterly since June 2009.³⁸ It includes both American and European, sovereign, and corporate issuers. I machine-read the data from these quarterly DTCC reports and remove all sovereign and all non-American reference entities. The DTCC reports contain some aggregate information

³⁷It introduces a timing error when a fund opens a new or closes an existing bond position. It should not be critical for my results because I compute the average number of mutual funds that hold the bond in active periods, which are quite prolonged by construction.

³⁸See [DTCC website](#).

on CDS transactions like the total number of clearing dealers and average daily notional amount. In this paper, I use only the fact that an entity is listed among 1000 most actively traded contracts and do not use additional characteristics provided by DTCC.

The reference entities in DTCC reports are text strings; other firm IDs are not provided. I match text strings from DTCC reports to issuer names from Mergent FISD database (after some usual text cleaning) by computing Jaro-Winkler distance and keeping all name pairs where the distance is less than 0.2. Then I manually check all matched pairs to ensure that I do not have any false matches. All the entities that were not matched or were not mentioned in the DTCC report in a given quarter are assigned the CDS dummy value of 0. All matched entities are assigned the value of 1 for all days in a given quarter. Among 1000 U.S. firms mentioned at least once in DTCC reports from 2010 to 2017, I match a bit more than 800. I might have some ‘true negatives’ in the final sample (the firms that were not matched due to some text processing errors), but it should not affect my results as long as ‘false positives’ (wrongly matched firms) are absent.

Appendix C. Additional Tables and Charts

	Mean	Median	S.D.	Min	5th	25th	75th	95th	Max	N.Obs.
Issue size, mln USD	655.24	500.00	708.38	0.61	9.40	250.00	850.00	2000.00	15000.00	5746678
Rating	7.97	7.33	3.27	1.00	4.00	6.00	10.00	14.00	21.00	5746678
Age, years	4.93	3.58	4.63	0.00	0.33	1.67	6.75	15.50	62.42	5746678
Maturity, years	9.37	6.50	8.05	1.00	1.50	3.50	12.08	27.33	29.92	5746678
Duration	6.75	5.57	4.49	0.84	1.41	3.20	9.00	15.86	27.93	5746678
Total return, %	0.03	0.03	1.25	-8.19	-1.85	-0.36	0.43	1.90	8.49	5746678
Credit spread, %	2.55	1.90	2.84	0.00	0.69	1.28	2.98	6.24	88.70	5746678
Average bid-ask, %	1.14	0.74	1.16	0.00	0.08	0.31	1.62	3.37	19.99	2308138
No. trades per day	6.45	3.00	11.17	1.00	1.00	2.00	7.00	22.00	2540.00	5746678
No. days since last trade	2.33	1.00	7.25	1.00	1.00	1.00	2.00	7.00	1436.00	5735632
C-to-C volume, % of size	0.50	0.00	1.97	0.00	0.00	0.00	0.08	2.50	15.99	5746678
C-to-D volume, % of size	0.01	0.00	3.52	-19.67	-4.35	-0.22	0.33	4.29	17.91	5746678
C-to-D volume , % of size	1.52	0.28	3.18	0.00	0.00	0.05	1.31	7.86	19.67	5746678

Table C1. Summary statistics of the unfiltered bond-day panel. This is a counterpart of Table I that shows how sample characteristics change in the full unfiltered bond-day panel (no restriction on the number of days since the previous trade).

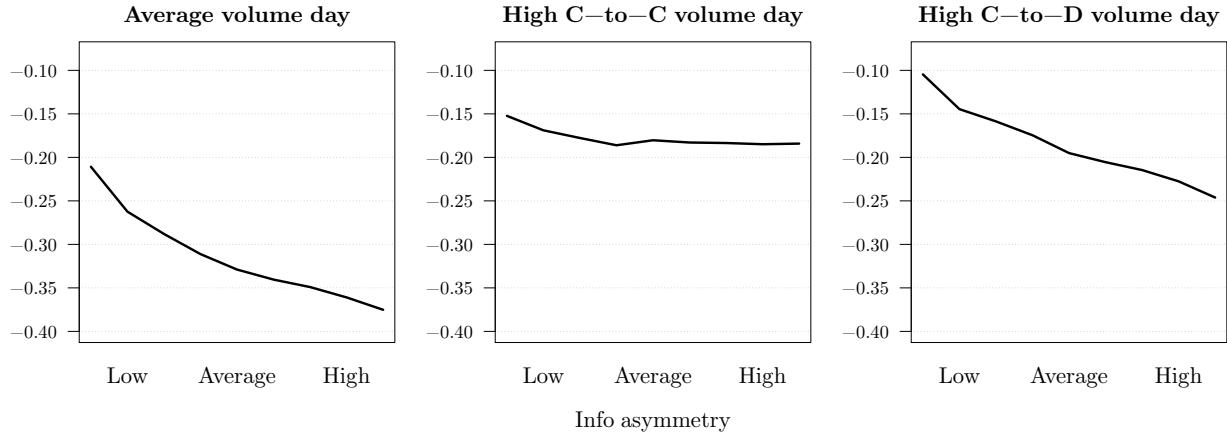


Figure C1. Point estimates for high-volume day reversals. The calculations are based on models (8) from Tables VI-VIII. On the x-axis from left to right are the deciles of information asymmetry proxies. For instance, ‘Low asymmetry’ bond is the one that has the number of fund owners, active CDS dummy, issue size, number of dealers, and issuer size all in the 90-th percentile and stock volatility in the 10-th percentile. ‘High asymmetry’ bond has the number of fund owners, active CDS dummy, issue size, number of dealers, and issuer size all in the 10-th percentile and stock volatility in the 90-th percentile. All other covariates from the regression models (average bid-ask spread, volume correlations, return volatility, and credit spread) are fixed at the median level. High C-to-C volume day is the day with C-to-C volume 2 standard deviations above the average (and average C-to-D volume); its reversal is $\mathbb{E}[\hat{\beta}_1|\text{covariates}] + 2\mathbb{E}[\hat{\beta}_2|\text{covariates}]$. High C-to-D volume day is the day with C-to-D volume 2 standard deviations above the average (and average C-to-C volume); its reversal is $\mathbb{E}[\hat{\beta}_1|\text{covariates}] + 2\mathbb{E}[\hat{\beta}_3|\text{covariates}]$. The reversal on the average volume day is simply $\mathbb{E}[\hat{\beta}_1|\text{covariates}]$.

	IG $\hat{\beta}_1$	HY	IG $\hat{\beta}_2$	HY	IG $\hat{\beta}_3$	HY
Intercept	-0.407*** (0.008)	-0.425*** (0.023)	0.108*** (0.011)	0.103*** (0.019)	0.107*** (0.008)	0.102*** (0.019)
Average bid-ask	-0.084*** (0.007)	-0.085*** (0.013)	0.018** (0.008)	-0.012 (0.014)	-0.038*** (0.006)	-0.050*** (0.012)
C-to-C vlm corr.	0.136*** (0.017)	0.124** (0.052)	0.092*** (0.019)	0.165*** (0.046)	-0.010 (0.016)	0.029 (0.046)
C-to-D vlm corr.	-0.088*** (0.016)	-0.024 (0.043)	-0.126*** (0.022)	-0.110** (0.050)	-0.025 (0.019)	0.004 (0.049)
No. funds	0.0002*** (0.0001)	0.0003* (0.0002)	-0.0002** (0.0001)	-0.0002* (0.0001)	0.0002** (0.0001)	0.00003 (0.0001)
CDS dummy	0.005* (0.003)	-0.006 (0.009)	-0.007* (0.004)	0.005 (0.008)	0.006** (0.003)	-0.009 (0.008)
Issue size	0.058*** (0.005)	0.089*** (0.019)	-0.009** (0.004)	-0.024* (0.014)	-0.001 (0.004)	0.013 (0.015)
No. dealers	0.001*** (0.0002)	0.002*** (0.001)	-0.001*** (0.0002)	0.0001 (0.0004)	-0.0002 (0.0002)	0.0003 (0.0005)
Issuer size	0.00005*** (0.00002)	0.00003 (0.0001)	0.00003 (0.00002)	-0.00000 (0.0001)	-0.0001*** (0.00002)	-0.0003*** (0.0001)
-Equity volatility	-0.541* (0.320)	1.119** (0.485)	-1.143*** (0.422)	0.492 (0.453)	0.544 (0.356)	-0.097 (0.407)
Risk controls	YES	YES	YES	YES	YES	YES
Observations	3,971	710	3,971	710	3,971	710
R ²	0.443	0.381	0.038	0.042	0.089	0.117

Note:

*p<0.1; **p<0.05; ***p<0.01

Table C2. Cross-sectional regressions of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$; investment-grade and high-yield bonds separately. Each model is an OLS regression with heteroscedasticity-consistent standard errors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. ‘No. funds’ is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. ‘No. dealers’ is the average number of unique dealers that intermediate trades in each bond. Stock return volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

	Mean	Med.	No.>0	No.<0	No.>0*	No.<0*	No. Obs.
$\hat{\beta}_1$	-0.3287	-0.3429	114	9649	0	8700	9763
$\hat{\beta}_2$	0.0712	0.0618	7079	2684	1671	185	9763
$\hat{\beta}_3$	0.0584	0.0564	6894	2869	2043	345	9763

Table C3. Summary statistics for the cross-section of volume-return coefficients (estimated controlling for the market return in the first-step regression). This is a counterpart of Table IV, but the first-step regression here is $R_{t+1} = \beta_0 + \beta_1 R_t + \beta_2 R_t \tilde{V}_t^{(c)} + \beta_3 R_t \tilde{V}_t^{(s)} + \beta_4 R_t^{\text{mkt}} + \epsilon_{t+1}$. The market return R_t^{mkt} is the return on Barclays IG Corporate Bond index.

	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_3$
Intercept	-0.438*** (0.005)	-0.448*** (0.007)	0.111*** (0.008)	0.099*** (0.009)	0.078*** (0.006)	0.089*** (0.007)
Average bid-ask	-0.044*** (0.006)	-0.052*** (0.005)	0.017*** (0.006)	0.017*** (0.007)	-0.046*** (0.005)	-0.042*** (0.005)
C-to-C vlm corr.	0.077*** (0.016)	0.074*** (0.016)	0.111*** (0.019)	0.101*** (0.019)	0.021 (0.016)	0.018 (0.017)
C-to-D vlm corr.	-0.059*** (0.014)	-0.069*** (0.014)	-0.119*** (0.020)	-0.120*** (0.021)	-0.013 (0.017)	-0.021 (0.018)
No. funds	0.001*** (0.0001)	0.0004*** (0.0001)	-0.0001** (0.0001)	-0.0001** (0.0001)	0.0003*** (0.0001)	0.0002*** (0.0001)
CDS dummy	-0.001 (0.003)	-0.002 (0.003)	-0.006* (0.003)	-0.006* (0.003)	0.004 (0.003)	0.002 (0.003)
Issue size	0.024*** (0.004)	0.028*** (0.004)	-0.012*** (0.004)	-0.011*** (0.004)	-0.009** (0.003)	-0.004 (0.004)
No. dealers	0.001*** (0.0002)	0.001*** (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0002)	-0.0004** (0.0002)	-0.0002 (0.0002)
Issuer size		-0.0001*** (0.00002)		0.00000 (0.00002)		-0.0001*** (0.00002)
-Equity volatility		-0.782** (0.323)		-0.974*** (0.340)		0.656** (0.279)
Risk controls	YES	YES	YES	YES	YES	YES
Observations	4,985	4,656	4,985	4,656	4,985	4,656
R ²	0.260	0.274	0.030	0.033	0.077	0.075

Note:

*p<0.1; **p<0.05; ***p<0.01

Table C4. Cross-sectional regressions of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ (market return included in the first-step regression). Each model is an OLS regression with heteroscedasticity-consistent standard errors. Volume-return coefficients are averaged for every bond across all active periods, so are the predictors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. ‘No. funds’ is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. ‘No. dealers’ is the average number of unique dealers that intermediate trades in each bond. Equity volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

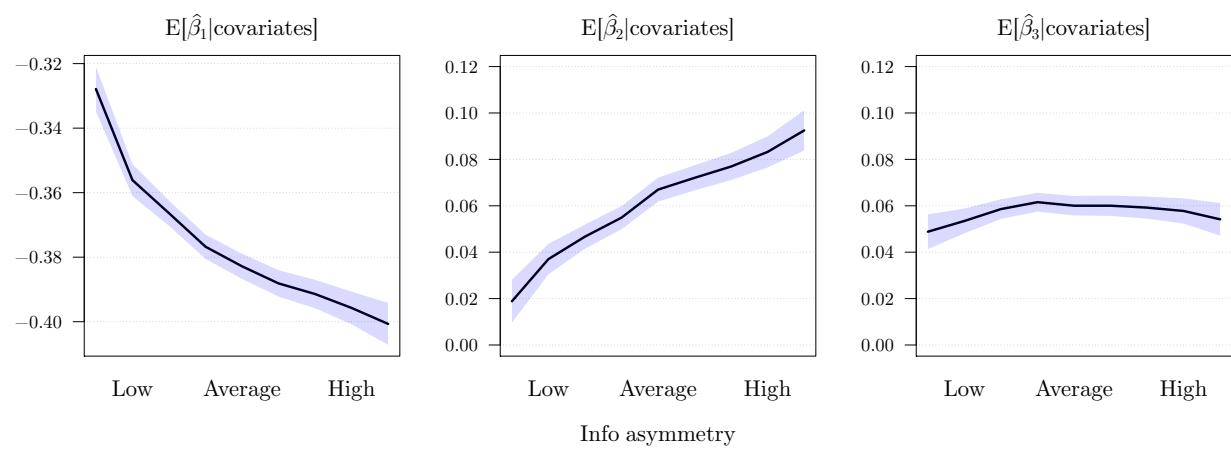


Figure C2. Point estimates and confidence intervals for the expected values of volume-return coefficients (market return included in the first-step regression). This figure is a counterpart of Figure 3, but the volume-return coefficients are estimated controlling for market return in the first-step regression (see Table C3).

	Mean	Med.	No.>0	No.<0	No.>0*	No.<0*	No. Obs.
$\hat{\beta}_1$	-0.3112	-0.3252	179	9644	12	8302	9823
$\hat{\beta}_2$	0.1126	0.0798	7414	2409	1948	157	9823
$\hat{\beta}_3$	0.0778	0.0732	7304	2519	2405	289	9823

Table C5. Summary statistics for the cross-section of volume-return coefficients (estimated controlling for volumes in the first-step regression). This is a counterpart of Table IV, but the first step equation here is $R_{t+1} = \beta_0 + \beta_1 R_t + \beta_2 R_t \tilde{V}_t^{(c)} + \beta_3 R_t \tilde{V}_t^{(s)} + \beta_4 \tilde{V}_t^{(c)} + \beta_5 \tilde{V}_t^{(s)} + \epsilon_{t+1}$.

	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_3$
Intercept	-0.395*** (0.007)	-0.398*** (0.007)	0.170*** (0.011)	0.164*** (0.013)	0.104*** (0.007)	0.117*** (0.008)
Average bid-ask	-0.091*** (0.006)	-0.103*** (0.006)	0.027*** (0.009)	0.027*** (0.010)	-0.055*** (0.006)	-0.050*** (0.006)
C-to-C vlm corr.	0.122*** (0.017)	0.119*** (0.017)	0.022 (0.026)	0.016 (0.026)	-0.010 (0.017)	-0.007 (0.017)
C-to-D vlm corr.	-0.074*** (0.016)	-0.075*** (0.017)	-0.099*** (0.037)	-0.072* (0.039)	0.019 (0.019)	0.007 (0.020)
No. funds	0.0003*** (0.0001)	0.0003*** (0.0001)	-0.0003*** (0.0001)	-0.0003*** (0.0001)	0.0003*** (0.0001)	0.0002*** (0.0001)
CDS dummy	0.002 (0.003)	0.002 (0.003)	-0.004 (0.005)	-0.004 (0.005)	0.005* (0.003)	0.002 (0.003)
Issue size	0.050*** (0.005)	0.046*** (0.005)	-0.016*** (0.006)	-0.014** (0.006)	-0.016*** (0.004)	-0.010** (0.004)
No. dealers	0.001*** (0.0002)	0.001*** (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0002)	-0.0003* (0.0002)	-0.0001 (0.0002)
Issuer size		0.00004** (0.00002)		-0.00003 (0.00003)		-0.0001*** (0.00002)
-Equity volatility		0.453 (0.286)		-0.330 (0.472)		0.645** (0.298)
Risk controls	YES	YES	YES	YES	YES	YES
Observations	5,018	4,691	5,018	4,691	5,018	4,691
R ²	0.356	0.368	0.029	0.027	0.076	0.076

Note: *p<0.1; **p<0.05; ***p<0.01

Table C6. Cross-sectional regressions of $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ (volumes included in the first-step regression). Each model is an OLS regression with heteroscedasticity-consistent standard errors. Volume-return coefficients are averaged for every bond across all active periods, so are the predictors. Average bid-ask is the percentage difference between the daily buy and sell prices, excluding inter-dealer trades. Volume correlations are the first autocorrelations of $\tilde{V}_t^{(c)}$ and $\tilde{V}_t^{(s)}$. ‘No. funds’ is the number of mutual funds that own the bond. CDS dummy equals 1 if the average Active CDS dummy for the bond across its active periods is above 0.5 and 0 otherwise. The issue size is the outstanding notional amount in bln USD. The issuer size is market cap in bln USD. ‘No. dealers’ is the average number of unique dealers that intermediate trades in each bond. Equity volatility is the average realized volatility of daily stock returns across all active periods for each bond. Risk controls include credit spread and realized bond return volatility.

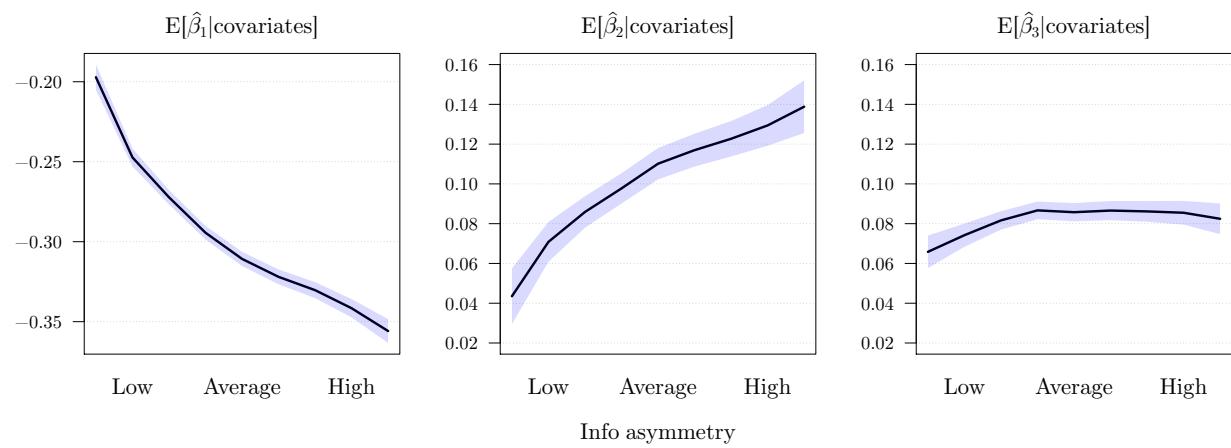


Figure C3. Point estimates and confidence intervals for the expected values of volume-return coefficients (volumes included in the first-step regression). This figure is a counterpart of Figure 3, but the volume-return coefficients are estimated controlling for trading volumes in the first-step regression (see Table C5).