Year : 2020

FROM MICROSTRUCTURE TO MACRO IMPLICATIONS: THREE ESSAYS ON CORPORATE BOND PRICING AND LIQUIDITY

Ivashchenko Alexey

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE FINANCE

FROM MICROSTRUCTURE TO MACRO IMPLICATIONS: THREE ESSAYS ON CORPORATE BOND PRICING AND LIQUIDITY

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Économiques, mention « Finance »

par

Alexey IVASHCHENKO

Directeur de thèse
Prof. Michael Rockinger

Jury

Prof. Felicitas Morhart, présidente
Prof. Johan Walden, expert interne
Prof. Pierre Collin-Dufresne, expert externe

LAUSANNE
2020

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE FINANCE

FROM MICROSTRUCTURE TO MACRO IMPLICATIONS: THREE ESSAYS ON CORPORATE BOND PRICING AND LIQUIDITY

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Économiques, mention « Finance »

par

Alexey IVASHCHENKO

Directeur de thèse
Prof. Michael Rockinger

Jury

Prof. Felicitas Morhart, présidente
Prof. Johan Walden, expert interne
Prof. Pierre Collin-Dufresne, expert externe

LAUSANNE
2020

Unil
UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

IMPRIMATUR

Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur Alexey IVASHCHENKO, titulaire d'un bachelor et d'un master en économie de l'Université d'État de Moscou, ainsi que d'un master en finance internationale de l'École des hautes études commerciales de Paris (HEC Paris), en vue de l'obtention du grade de docteur ès Sciences économiques, mention « finance ».

La thèse est intitulée :

**FROM MICROSTRUCTURE TO MACRO IMPLICATIONS:
THREE ESSAYS ON CORPORATE BOND PRICING AND LIQUIDITY**

Lausanne, le 19 mai 2020

Le doyen

Jean-Philippe Bonardi

HEC Lausanne
Le Décanat
Tél. +41 21 692 33 40 | Fax +41 21 692 33 05
www.hec.unil.ch | hecdoyen@unil.ch

Members of the thesis committee

Prof. Michael Rockinger

Professor of Finance, HEC Lausanne, University of Lausanne

Thesis supervisor

Prof. Johan Walden

Professor of Finance, HEC Lausanne, University of Lausanne

Internal member of the thesis committee

Prof. Pierre Collin-Dufresne

Professor of Finance, Ecole Polytechnique Fédérale de Lausanne (EPFL)

External member of the thesis committee

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Alexey IVASHCHENKO

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________ Date: May 15, 2020

Prof. Michael ROCKINGER
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Alexey IVASHCHENKO

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ____________________ Date: May 18, 2020

Prof. Pierre COLLIN-DUFRESNE
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Economics,
Subject area Finance

I hereby certify that I have examined the doctoral thesis of

Alexey IVASHCHENKO

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ____________________ Date: 13/5 2020

Prof. Johan WALDEN
Internal member of the doctoral committee

Acknowledgements

My doctoral thesis comprises three papers I have been working on as a doctoral student at the University of Lausanne and the Swiss Finance Institute in 2014–2020. The papers investigate different aspects of corporate bond pricing and liquidity: from microstructure-related peculiarities of trading activity and revelation of private information to the implications for aggregate credit risk premia estimation and macroeconomic forecasting. The first chapter of this thesis, 'Corporate Bond Price Reversals', was my job market paper on the 2019–2020 academic job market that allowed me to join VU Amsterdam as a tenure-track Assistant Professor of Finance after the completion of this thesis.

There has to be only my name on the title page of this thesis, but it took more than one man's effort to complete this journey. First and foremost, I would like to thank my thesis advisor Michael Rockinger for being there for me during my doctoral times. He claims not to write papers with his Ph.D. students. That is true; he does more than that. One would never get as much attention, advice, and motivation from a coauthor as Michael's students receive. I can not imagine a better academic mentor than Michael. Together with Michael, I would like to thank my dissertation committee members Pierre Collin-Dufresne and Johan Walden, for their advice, time, and incredible support on the academic job market. To have an opportunity to talk to Pierre and Johan about my papers was a privilege that made me a better researcher. I also thank my co-author Artem Neklyudov for his contribution to the third chapter of this thesis, which is our joint work.

The doctoral school has been a joyful experience thanks to my classmates Jakub Hajda, Sander Willems, and Julien Blatt. If there's a chance to become colleagues again sometime in the future, I'll take it hands down. That extends to the entire finance faculty and graduate schools at UNIL, EPFL, and the Swiss Finance Institute. If I had to choose a doctoral school again, I would not think twice.

Last but not least, I would like to thank my beloved family and friends for always being supportive and understanding. My wife Anna and my parents Elena and Sergey deserve a medal for getting my back on non-academic matters. My first kid, Pavel, was born in 2018 and has been a source of inspiration and joy ever since. His contribution to the completion of my doctoral degree cannot be overestimated.

Contents

1 Corporate Bond Price Reversals **6**

1.1 Introduction . . . 6

1.2 Data and measurements . . . 14

1.2.1 Data sources . . . 14

1.2.2 Sample filtering and 'active periods' . . . 14

1.2.3 Volume measures . . . 16

1.2.4 Proxies for information asymmetry . . . 18

1.3 Volume-return relationship . . . 20

1.4 Determinants of volume-return coefficients . . . 22

1.4.1 Empirical design . . . 22

1.4.2 Main results . . . 26

1.4.3 Further evidence . . . 33

1.5 Implications for investment strategies . . . 36

1.6 The model . . . 39

1.6.1 The economy . . . 40

1.6.2 Model equilibrium . . . 43

1.6.3 Trading volume in the model . . . 44

1.6.4 Volume-return relationship and information asymmetry . . . 45

1.7 Conclusion . . . 49

2 Credit Spreads, Daily Business Cycle, and Corporate Bond Returns Predictability **52**

2.1 Introduction . . . 52

2.2 Sample Characteristics . . . 55

2.3 Measuring Excess Bond Premium . . . 58

2.4 Forecasting the Business Cycle . . . 65

2.5 Forecasting Corporate Bond Returns . . . 70

2.5.1 EBP and Excess Corporate Bond Market Returns . . . 71

2.5.2 'Real-time' EBP as a Predictor of Market Returns . . . 75
2.5.3 EBP and Other Corporate Bond Risk Factors . . . 79
2.5.4 Corporate Bond Market-timing Strategy . . . 84
2.6 Conclusion . . . 88

3 (In)frequently Traded Corporate Bonds 90
3.1 Introduction . . . 90
3.2 Data and measurements . . . 94
3.3 (In)frequently traded bonds . . . 99
3.3.1 Main characteristics . . . 99
3.3.2 Trading volume and frequency . . . 101
3.3.3 Mutual fund holdings, trading frequency, and illiquidity . . . 104
3.3.4 Changes in mutual fund demand and trading frequency . . . 105
3.4 Trading frequency and public information about bond issuers and issues . . . 107
3.5 Trading frequency and returns . . . 114
3.6 Concluding Remarks . . . 119

Bibliography 122

Appendices 128

A Appendix of Chapter 1 128
A.1 Aspects of the model . . . 128
A.1.1 Log-linear approximation of returns . . . 128
A.1.2 Learning by uninformed investors . . . 129
A.1.3 Optimal demands . . . 130
A.1.4 Existence of the equilibrium . . . 131
A.1.5 Derivation of the volume-return relationship . . . 134
A.2 Data and sample . . . 137
A.2.1 Sample selection . . . 137
A.2.2 Actively traded CDS contracts . . . 138
A.2.3 Winsorization . . . 138
A.3 Additional Tables and Charts . . . 139

B Appendix of Chapter 2 146
B.1 Constructing the Dataset . . . 146
B.1.1 Data Sources . . . 146
B.1.2 Daily Measure of Merton's Distance-to-default . . . 149

B.2 Explanatory Variables 150
B.3 Time fixed effects in regressions for spreads 152
B.4 EBP and VIX as predictors of returns 156

C Appendix of Chapter 3 **158**
C.1 Sample selection 158
C.2 SEC N-Q forms and holdings data 159
C.3 Additional tables and charts 162

Chapter 1

Corporate Bond Price Reversals

1.1 Introduction

Sophisticated investors used to own a substantial fraction of U.S. corporate bonds around the global financial crisis of 2008–2009. Figure 1.1 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]

In this chapter, 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

[1] Business 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 2013a). Berndt and Zhu (2018) provide a model that links higher dealer inventory costs with lower market efficiency post-crisis.

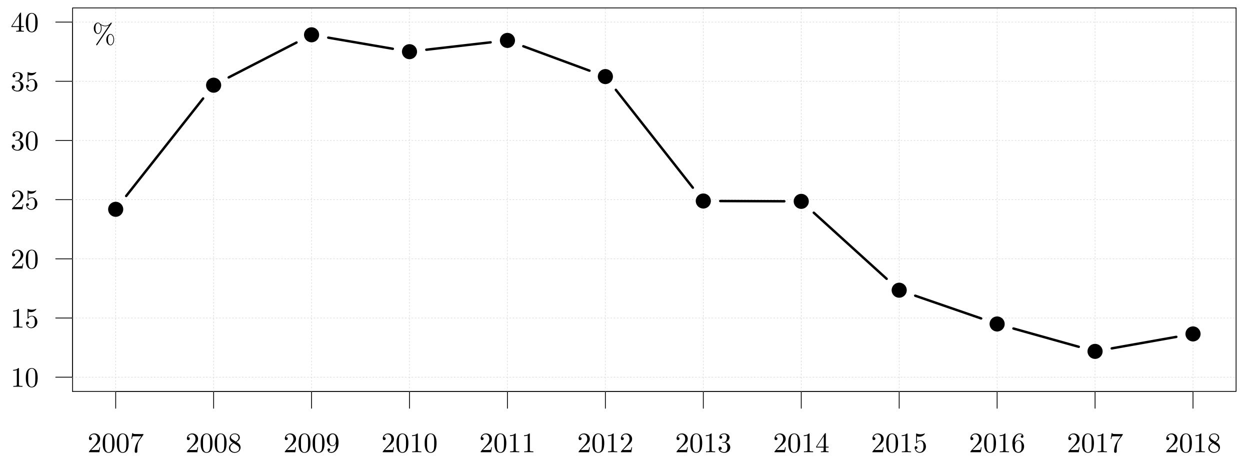


Figure 1.1. Hedge funds' corporate bond holdings, in % of the combined holdings of insurance companies, pension funds, mutual funds, and ETFs. I use U.S. Flow of Funds (FF) data to calculate the ratio. The FF data do not separate hedge funds, but the industry tradition is to interpret households' corporate bond holdings as the ones dominated by hedge funds.

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 chapter 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.

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

[2]The 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.

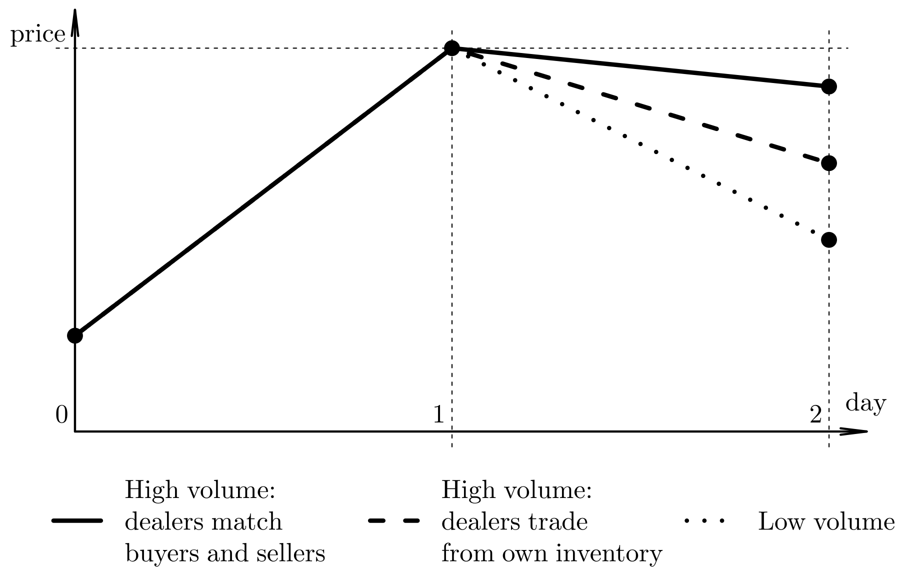


Figure 1.2. Stylized price reversal paths for a high-information-asymmetry bond. On day 1, the trading volume is either low or high. The solid line shows a reversal path on a high-volume day when dealers' end-of-day inventory (in this particular bond) does not change, and dealers buy from some investors 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.

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 1.2 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-

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

return relationship for individual corporate bonds:[4]

$$R_{t+1} = \beta_0 + \underbrace{\left(\beta_1 + \beta_2 \cdot \text{Inventory-neutral volume}_t + \beta_3 \cdot |\Delta\text{Inventory}|_t\right)}_{\text{Return autocorrelation}} R_t + \epsilon_{t+1}, \quad (1.1)$$

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 (1.1), β_1 measures the reversal on a low-volume day, while β_2 and β_3 capture how the reversal changes following high-volume days with different dealers' capital commitment. The volume-return relationship (1.1) 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,

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

[5] In this chapter, 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 (1.1) are zero.

$\hat{\beta}_2$ 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 chapter 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 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 work contributes to several streams of corporate bond literature. The chapter 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

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

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.

My work also contributes to the debate on information-driven trading in the corporate bond market. Asquith et al. (2013a) 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 analysis, 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 work 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.

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

The latter finding links this chapter 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 work 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 (1.1) 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'

[8] I consider aggregate dealers' corporate bond inventory in this chapter and do not investigate end-of-day inventory changes of individual dealers.

[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.

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 results suggest that this policy change will increase information asymmetry between investors in bonds included in the pilot. Higher asymmetry 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 chapter is organized as follows. Section 1.2 talks about the bond sample and the steps I take to estimate a volume-return relationship for individual bonds. Section 1.3 presents estimated volume-return coefficients, and Section 1.4 investigates its determinants, in particular, information asymmetry proxies, in a cross-section of bonds. Section 1.5 discusses the implications of my results for reversal investment strategies. Section 1.6 solves a stylized theoretical model of competitive corporate bond trading and discusses a volume-return relationship an econometrician observing such an economy recovers. Section 1.7 concludes.

[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.

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

1.2 Data and measurements

1.2.1 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 A.2.

1.2.2 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 (1.1), 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 market indices, which also drives substantial institutional rebalancing and creates abnormal price patterns that are not the primary focus of this study.

Table 1.1 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]

[12] Here 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.

[13] I present the same summary statistics for the full, unfiltered bond-day panel in Table A.1 in Appendix A.3. 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.

	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 1.1. 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 A.2. The same summary statistics for a full, unfiltered bond-day panel is in Table A.1 in Appendix A.3.

1.2.3 Volume measures

To construct a proxy for the inventory-neutral trading volume of equation (1.1), 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':

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

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

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 (1.1), I consider the absolute value of the C-to-D trading volume, $|V_{it}^{(s)}|$. Table 1.1 shows that the absolute value of the C-to-D volume is on average several times higher than the C-to-C volume.

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

Table 1.2. 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.

Table 1.2 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 $\text{Corr}\left(V_t^{(c)}, V_t^{(s)}\right) = 0$, i.e., bond inventory is equally likely 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 1.2.

1.2.4 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 (1.1), 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 (1.1) 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 1.3 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 1.3. 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 this chapter. 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 (1.1), is similar to the expected effect of information asymmetry. The impact of the bid-ask bounce on β_2 and β_3 in equation (1.1) 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.

1.3 Volume-return relationship

I estimate equation (1.1) 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:

$$R_{t+1} = \beta_0 + \beta_1 R_t + \beta_2 R_t \tilde{V}_t^{(c)} + \beta_3 R_t \tilde{V}_t^{(s)} + \epsilon_{t+1}. \tag{1.2}$$

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, standardized[15] for every active period separately, and $\tilde{V}_t^{(s)}$ is *the absolute value* of

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

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.

	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 1.4. Summary statistics of the estimated volume-return coefficients of equation (1.2). 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.

Table 1.4 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, 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-forth 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.

1.4 Determinants of volume-return coefficients

1.4.1 Empirical design

In the introduction, I put forward an intuition on how volume-return coefficients β_1, β_2, and β_3 in equation (1.2) 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 1.6. 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 $i \in \{1, \ldots, N\}$ where N is the total number of bonds). I fit the following model for each n (i.e., each volume-return coefficient) separately:

$$\begin{aligned}\hat{\beta}_{n,i} = c_{n,1}\, &\underbrace{(\text{No. funds, CDS, Issue/issuer size, No. dealers, –Equity volatility})_i}_{\text{Info asymmetry proxies}} + \\ &+ c_{n,2}\, \underbrace{(\text{Bid-ask, C-to-C/D volume correlation, Bond volatility, Credit spread})_i}_{\text{Controls}} + \\ &+ c_{n,0} + \epsilon_{n,i}, \end{aligned} \tag{1.3}$$

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 (1.3): 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.

[16] active 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).

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 1.4 that returns tend to continue following high C-to-C and C-to-D volume days (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

[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 work. I obtained similar results with alternative bond illiquidity measures as well (Amihud, Roll, price inter-quartile range).

[18] This 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 (2007a). This may also explain why the average autocorrelation of $\tilde{V}_t^{(c)}$ is relatively low in the data (see Table 1.2).

explain some differences between estimated volume-return coefficients. Same happens in the theoretical model of Section 1.6. 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.

	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 1.5. 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 A.2.

Table 1.5 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 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 (1.3) as Table 1.5 shows.

1.4.2 Main results

Tables 1.6–1.8 present estimated regressions (1.3) of volume-return coefficients on information asymmetry proxies and controls. Table 1.6 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 volatility – 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 1.6 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 1.6, 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 1.6, 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 1.7 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 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 1.7 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 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.

	Dependent variable: $\hat{\beta}_1$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	-0.331^{***}	-0.301^{***}	-0.416^{***}	-0.399^{***}	-0.349^{***}	-0.301^{***}	-0.429^{***}	-0.450^{***}
	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.008)
Average bid-ask	-0.055^{***}	-0.062^{***}	-0.054^{***}	-0.098^{***}	-0.070^{***}	-0.067^{***}	-0.064^{***}	-0.073^{***}
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)
No. funds	0.033^{***}						0.007^{***}	0.007^{***}
	(0.002)						(0.002)	(0.002)
CDS dummy		0.003^{*}					0.002	0.001
		(0.001)					(0.001)	(0.001)
Issue size			0.059^{***}				0.046^{***}	0.040^{***}
			(0.003)				(0.004)	(0.004)
No. dealers				0.044^{***}			0.013^{***}	0.017^{***}
				(0.002)			(0.003)	(0.003)
Issuer size					0.024^{***}			0.011^{***}
					(0.002)			(0.002)
–Equity volatility						0.0001		0.005^{**}
						(0.002)		(0.002)
Risk controls	YES	YES	YES	YES	YES	YES	YES	YES
Vlm 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	0.247	0.391	0.331	0.284	0.255	0.398	0.417
Note:	*p<0.1; **p<0.05; ***p<0.01							

Table 1.6. 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.

Also, observe in Table 1.7 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 1.7 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 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 1.8 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

	Dependent variable: $\hat{\beta}_2$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.090***	0.082***	0.113***	0.117***	0.088***	0.076***	0.125***	0.126***
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.006)	(0.008)	(0.009)
Average bid-ask	0.001	0.003	0.001	0.017***	0.006	0.005	0.008*	0.010*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
No. funds	−0.012***						−0.004**	−0.003*
	(0.002)						(0.002)	(0.002)
CDS dummy		−0.004**					−0.003*	−0.003*
		(0.002)					(0.002)	(0.002)
Issue size			−0.017***				−0.009***	−0.010***
			(0.002)				(0.003)	(0.003)
No. dealers				−0.017***			−0.009***	−0.010***
				(0.002)			(0.003)	(0.003)
Issuer size					−0.005***			−0.0002
					(0.002)			(0.002)
–Equity volatility						−0.003		−0.005**
						(0.002)		(0.002)
Risk controls	YES	YES	YES	YES	YES	YES	YES	YES
Vlm 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	0.013	0.026	0.025	0.015	0.014	0.030	0.036
Note:	*p<0.1; **p<0.05; ***p<0.01							

Table 1.7. 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.

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 1.8 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

	Dependent variable: $\hat{\beta}_3$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.041***	0.041***	0.046***	0.051***	0.048***	0.042***	0.050***	0.054***
	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)
Average bid-ask	−0.046***	−0.046***	−0.047***	−0.044***	−0.043***	−0.042***	−0.041***	−0.038***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
No. funds	0.003**						0.005***	0.003
	(0.001)						(0.002)	(0.002)
CDS dummy		0.002*					0.002*	0.001
		(0.001)					(0.001)	(0.001)
Issue size			−0.001				−0.001	0.0001
			(0.002)				(0.002)	(0.002)
No. dealers				−0.003*			−0.005**	−0.003
				(0.002)			(0.002)	(0.002)
Issuer size					−0.005***			−0.005***
					(0.002)			(0.002)
–Equity volatility						0.003		0.003
						(0.002)		(0.002)
Risk controls	YES	YES	YES	YES	YES	YES	YES	YES
Vlm 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	0.106	0.106	0.106	0.106	0.105	0.103	0.108	0.106
Note:	*p<0.1; **p<0.05; ***p<0.01							

Table 1.8. 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.

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 1.6–1.8 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 1.6–1.8 (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

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

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.

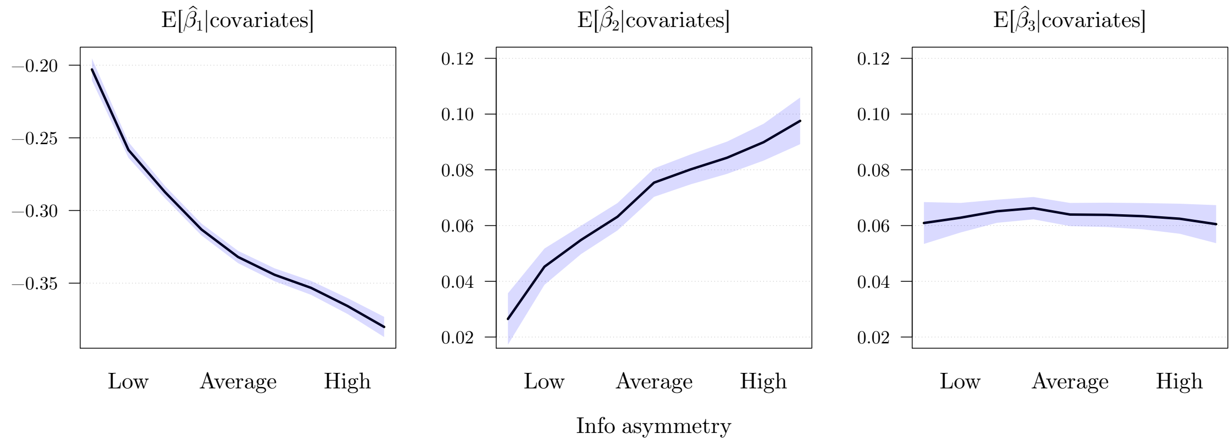


Figure 1.3. Point estimates and confidence intervals for the expected values of volume-return coefficients. The calculations are based on models (8) from Tables 1.6-1.8. On the x-axes from left to right are the deciles of information asymmetry proxies. For instance, 'Low asymmetry' bond is the one that 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.

Figure 1.3 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 1.3 shows an additional impact of high C-to-C volumes on next-day reversals. The average values of $\hat{\beta}_2$ are monotonically increasing 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 A.1 in Appendix A.3 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 1.3 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 1.3 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.

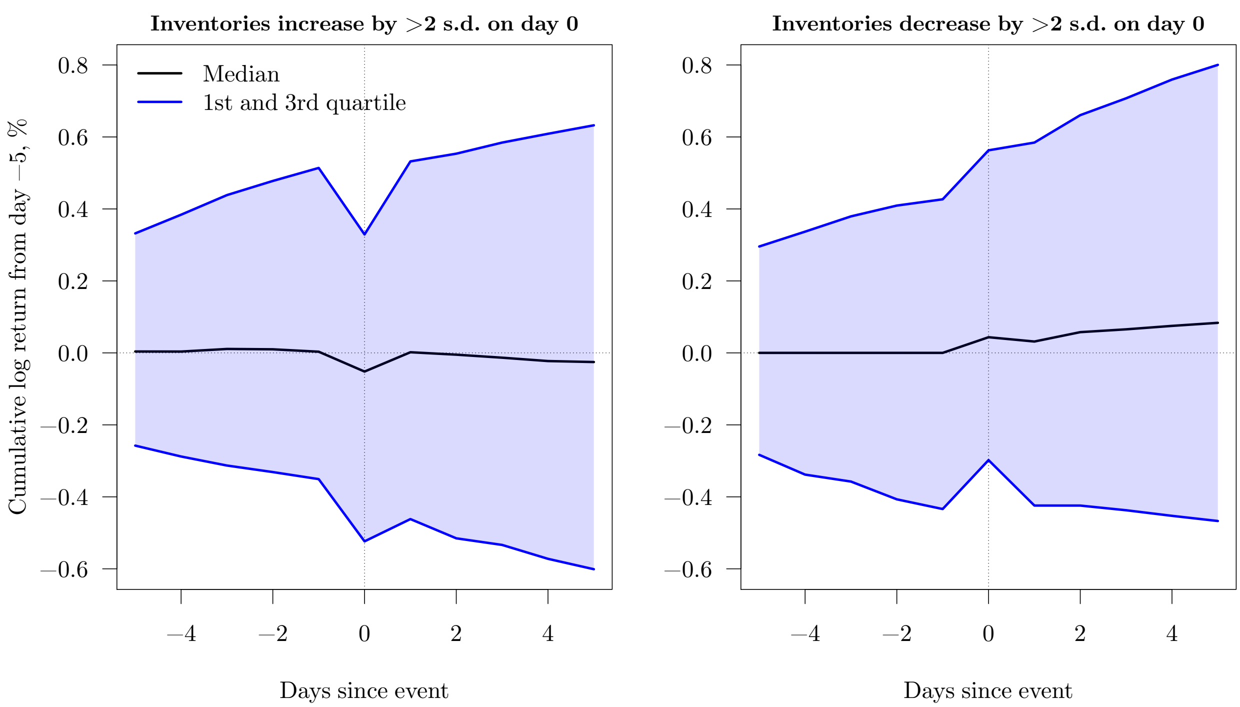


Figure 1.4. Cumulative returns around days with large bond inventory changes. The 'event' that happens on day 0: broker-dealers bond inventory increases or decreases by more than 2 standard deviations (computed per bond per active period) *and* it is the only type of trading that occurs on day 0 (inventory 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 1.2, such days correspond to high C-to-D volume and *zero* C-to-C volume. Figure 1.4 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 1.4 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 pronounces 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 1.4 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.

1.4.3 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 1.9, I present the estimates of model (1.3) 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 1.9 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 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.

In Appendix A.3, I present further empirical results. Table A.2 estimates equation (1.3) 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 A.2 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 A.2 confirms the results that have been established in the pooled sample.

I also consider alternative specifications of equation (1.2) to address the omitted variable problem that may render the estimates of volume-return coefficients biased. In Ap-

[20] I 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.

pendix A.3, 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 (1.2). Tables A.3 and A.5 present summary statistics of volume-return coefficients for these two cases, while Tables A.4 and A.6 show the dependence on information asymmetry proxies. Figures A.2 and A.3, the counterparts of Figure 1.3, 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.

	$\hat{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_3$
Average bid-ask	-0.066^{***}	-0.075^{***}	0.005	0.007	-0.006	-0.004
	(0.010)	(0.010)	(0.011)	(0.010)	(0.008)	(0.008)
No. funds	0.009^{***}	0.010^{***}	-0.008^{**}	-0.008^{**}	0.002	0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
CDS dummy	0.012	-0.004	0.001	-0.003	0.001	-0.005
	(0.010)	(0.009)	(0.010)	(0.010)	(0.008)	(0.008)
Issue size	0.029^{***}	0.023^{***}	-0.00003	-0.001	-0.002	-0.002
	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)
No. dealers	0.016^{***}	0.026^{***}	-0.011^{***}	-0.013^{***}	-0.011^{***}	-0.009^{**}
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Issuer size		0.044^{***}		0.008		-0.009
		(0.008)		(0.010)		(0.009)
–Equity volatility		0.026^{***}		-0.013^{*}		0.023^{***}
		(0.006)		(0.007)		(0.006)
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	0.553	0.568	0.115	0.131	0.217	0.204
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table 1.9. 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.

1.5 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 bondholders 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 pre-

[21] I 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.

[22] For 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.

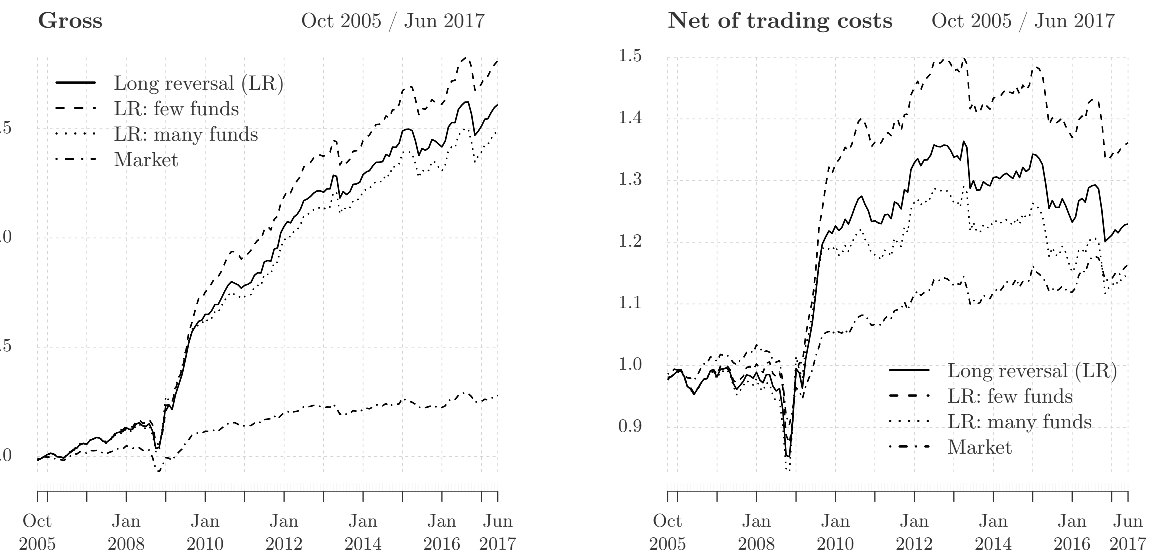
vious section suggest that in-sample and following average-volume periods the reversals are stronger for bonds with more information asymmetry. So, one might expect the reversal portfolio with more information asymmetry to outperform the reversal portfolio with less information asymmetry out-of-sample .

	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 1.10. 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 1.1).

Table 1.10 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 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.



(a) Not adjusted for trading costs

(b) Adjusted for trading costs

Figure 1.5. 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 1.1).

Figure 1.5 shows how reversal returns accumulate over time. Observe in Figure 1.5a 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 1.5b plots portfolio values net of trading costs and tells a similar story except the 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 different information asymmetry signals and potentially improve the performance of the reversal strategy on corporate bonds.

1.6 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 (1.2), which I estimate in the empirical part of the chapter, 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 chapter.

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 1.1) 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 1.2 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 (1.2) and provide additional implications for my empirical analysis compared to the baseline model of Llorente et al. (2002).

1.6.1 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

[23]To 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.

(1988) decomposition applied to such a bond yields the log return r_t of the following form:

$$r_{t+1} \approx \kappa + c(1-\theta) + \theta p_{t+1} - p_t - d_{t+1}, \tag{1.4}$$

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.

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 $\{d, p, n, f, s, g, z\}_{0,\ldots,t}$ while the information set of the uninformed investors is $\{d, p, n, f, s\}_{0,\ldots,t}$.

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

[24] For the derivation see Appendix A.1.1.

[25] Here 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.

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

$$s_{t+1} = \delta s_t + \epsilon_{t+1}, \tag{1.5}$$

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

$$\begin{aligned} W_{t+1}^{(1)} &\approx W_t^{(1)} + X_t^{(1)} r_{t+1} + z_t(1 + n_{t+1}), \\ W_{t+1}^{(2)} &\approx W_t^{(2)} + X_t^{(2)} r_{t+1}. \end{aligned}$$

The model setup is different from Llorente et al. (2002) in two ways. First, I work with log returns approximated in (1.4) 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 (1.5) 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

[26] The 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.

[27] As 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 corporate bonds purchased by clients matched by client sales in a given period and net changes in broker-dealer inventory.

1.6.2 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:

$$\tilde{p}_t = -a(g_t + bz_t + es_t). \tag{1.6}$$

Observe that the steady-state level of log bond price is 0 as in the linear approximation of log return (1.4).

Given the pricing function (1.6), the equation for returns (1.4) re-writes as:[28]

$$r_{t+1} = -\theta\left(f_{t+1} - m_f\right) + \theta\tilde{p}_{t+1} - \tilde{p}_t - g_t. \tag{1.7}$$

The expression for conditional expected returns follows from (1.7):

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

The informed investors know g_t, hence $\mathbb{E}_t^{(1)}\left[g_t\right] = g_t$. The uninformed investors observe $\tilde{p}_t$ and s_t and estimate $\mathbb{E}_t^{(2)}\left[g_t|\tilde{p}_t, s_t\right]$. I show in Appendix A.1.2 that

$$\mathbb{E}_t^{(2)}\left[g_t|\tilde{p}_t, s_t\right] = \gamma(g_t + bz_t), \tag{1.8}$$

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.

[28] In what follows, I replace an approximate equality in (1.4) with the exact one.

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.1.4 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.

1.6.3 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

$$\left|V_{c,t}^{(1)}(\Delta g_t, \Delta z_t)\right| = \left|V_{c,t}^{(2)}(\Delta g_t, \Delta z_t)\right| = \left|\alpha\left(\Delta g_t + \sigma_{rn}\Delta z_t\right)\right|, \tag{1.9}$$

[29]See Appendix A.1.3.

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 $v_{c,t} = |\alpha(\Delta g_t + \sigma_{rn}\Delta z_t)| \geqslant 0$. 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 1.3 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.

1.6.4 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

$$\mathbb{E}_t\left[r_{t+1}|r_t, v_{c,t}, v_{s,t}\right] \approx \left(\beta_1 + \beta_2 v_{c,t}^2 + \beta_3 v_{s,t}^2\right) r_t, \tag{1.10}$$

[30] All 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.1.5. This volume-return relationship is a theoretical counterpart of equation (1.2) estimated in the empirical part of the chapter. Unlike equation (1.2), equation (1.10) 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 (1.2) 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 1.6, 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

[31] Since an econometrician knows the sign of inventory changes, she could write an analog of equation (1.10) 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 (1.10) on the absolute value of inventory changes.

of returns of 1%.[32] The latter stays fixed in all numerical solutions; this is an additional constraint I impose on the solutions of the model.[33] Figure 1.6 represents the cross-section of bonds with the same unconditional risk but different contributions of public, private, and liquidity shocks to return variance.

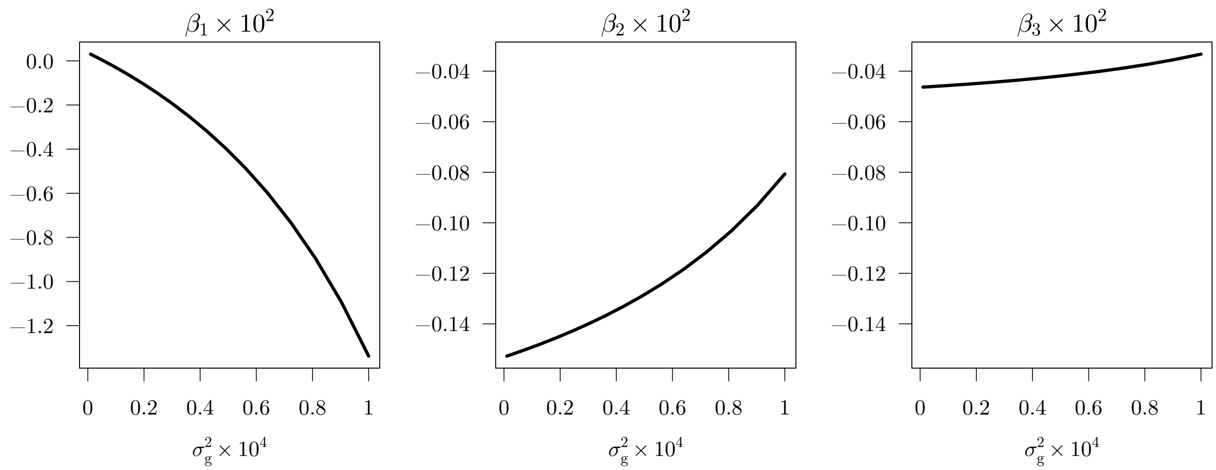


Figure 1.6. Dependence of β_1, β_2, and β_3 on information asymmetry σ_g^2 holding total return variance fixed. Each point on the curves is a numerical solution of the model. I obtain the relationships between σ_g^2 and β coefficients by varying σ_g from 0 to 1% holding an unconditional standard deviation 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 1.6 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

[32]In Figure 1.6, 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 (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.

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 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 1.6 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 1.6 is due to equilibrium e (price impact of inventory-changing trades) changing with σ_g.

The shape of the lines in Figure 1.6 matches closely the shape of their empirical counterparts presented in Figure 1.3. 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

[34] Here 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 (1.7) and (1.8) one finds that r_t is negative when $\frac{a}{\gamma}\mathbb{E}^{(2)}_{t-1}[g_{t-1}] < g_{t-1}$ 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.

is not the part of the benchmark model). In reality, as Section 1.3 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 chapter. As the extent of informed trading increases, returns following high-volume days are less likely to revert, especially when dealers are not trading from their inventory capacity.

1.7 Conclusion

In this chapter, 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.

Chapter 2

Credit Spreads, Daily Business Cycle, and Corporate Bond Returns Predictability

2.1 Introduction

Credit spreads forecast economic activity. Gilchrist and Zakrajšek (2012) elaborated on this statement; there is a particular portion of credit spreads that is of most importance for activity forecasts. It is a part of the spread that is *not* explained by corporate credit risk, called the credit risk premium or the excess bond premium (EBP). The first part of this chapter shows what stands behind the forecasting power of the EBP.

I argue that the forecasting power of the EBP hinges on the information about aggregate business risk and bond liquidity risk contained in credit spreads and show how to extract this information using daily frequency. I construct a large bond-day panel of credit spreads from transactions recorded in TRACE and measure corporate credit risk, bond-specific liquidity risk, and aggregate business risk at the daily frequency. When I project spreads on corporate credit risk only (as in Gilchrist and Zakrajšek, 2012), I confirm that the residual forecasts future economic activity. However, when I further project spreads on aggregate business risk as measured by the Aruoba-Diebold-Scotti daily business conditions index (ADS index) and bond liquidity risk, as measured by the Amihud measure, the forecasting power of the residual portion of spreads for macroeconomic variables largely goes away.[1]

Following finance literature, I interpret the residual portion of credit spread unexplained by corporate credit risk, bond liquidity risk, and aggregate business risk as the *credit risk premium*. The second part of this chapter demonstrates that my measure of the credit risk premium is a forecast of corporate bond market returns. The forecasting power is absent when one considers instead the residuals from the projection of spreads on corporate credit risk only. This result is robust to different estimation windows and different bond market portfolios. Moreover, the risk premium forecasts returns even when it is estimated in real time with the information available only on the estimation date.

I remain agnostic about what this return-forecasting component of credit spreads is. Yet, I demonstrate what it is surely not. My work shows that neither bond pricing factors of Bai et al. (2019), including contemporaneous bond market returns per se, nor stock market factors can explain the time series variation of my credit risk premium measure. The models with my credit risk premium on the right-hand side, in addition to other bond pricing factors, however, forecast returns on diverse size, maturity, and industry corporate bond portfolios better than the models without it. This result is robust to exclusion of the subprime crisis episode from the sample.

I exploit the forecasting power of the risk premium to construct a corporate bond market-timing strategy that delivers risk-return characteristics superior to the buy-and-hold market strategy. My strategy assumes weekly portfolio rebalancing and uses only one risky instrument, an investable aggregate corporate bond market index, which is bought and sold depending on predicted corporate bond market excess returns. On a testing sample, my pre-

[1] The ADS index does not contain any bond or stock market indicators as inputs.

dictive model successfully forecasts market returns out-of-sample, and the strategy delivers total return and a Sharpe ratio 1.5 times higher than the corporate bond market index.

The first part of this chapter on macro forecasting properties of the EBP feeds into several discussions in the literature. From the perspective of EBP estimation and predictive power, this work is related to the work by Gilchrist and Zakrajšek (2012), De Santis (2017), and Nozawa (2017). In particular, De Santis (2017) constructed a monthly credit risk premium free from aggregate business risk on European multi-country data but reached a different conclusion regarding its forecasting properties. From the perspective of empirical credit spread modeling, this chapter contributes to the 'credit spread puzzle' literature stemming from Collin-Dufresne, Goldstein, and Martin (2001). I demonstrate that aggregate business risk, as measured by the daily business cycle index, is able to explain a significant portion of common variation in credit spreads at the daily frequency. In this respect, this chapter is related to the results by d'Avernas (2017), who estimates a joint structural model of credit spreads and equity volatility to argue that firms' time-varying aggregate asset volatility helps to explain both the dynamics of credit spreads and their forecasting power for economic activity. There are no direct references for the second part of this chapter that investigates asset pricing properties of the EBP. To the best of my knowledge, this is the first study to establish the forecasting power of EBP for corporate bond market returns.

The chapter is organized as follows. Section 2.2 discusses the data sample. Section 2.3 estimates the credit risk premium by fitting alternative models to the bond-day panel of credit spreads. Forecasting power of the risk premium for macroeconomic activity is discussed in Section 2.4. Section 2.5 shows that the EBP forecasts excess bond market returns, does multiple robustness tests, and presents an investment strategy to benefit from the forecasting power of the risk premium. Section 2.6 concludes the work.

2.2 Sample Characteristics

I merge daily bond trades from TRACE with bond characteristics from Mergent Fixed Income Securities Database (FISD) and issuing firm characteristics from Compustat and CRSP for senior unsecured corporate bonds with fixed coupon schedules. My data construction approach is presented in detail in Appendix B.1. The constructed sample is an unbalanced bond-day panel with around 2 million bond-day observations that span a period from Oct 2004 to Dec 2014. The number of bonds sampled per day is, on average, 823 with a standard deviation of 111. Summary statistics for the panel are presented in Table 2.1.

An average bond in the sample has been issued approximately six years ago and has about nine years to maturity. It has an outstanding amount of about 600 million USD and pays a coupon of 6%. It is an investment-grade security rated between BBB+ and BBB and is traded six times per day. Its yield to maturity is about 5%, approximately 2.4% above its risk-free counterpart. The latter number is the credit spread measure constructed following Gilchrist and Zakrajšek (2012). I call it either the GZ spread or simply the spread.

To control for illiquidity, I use a daily Amihud measure AMH_t computed for each bond for each day t when the bond was traded:

$$AMH_t = \frac{1}{N_t} \sum_{j=1}^{N_t} \frac{|r_{t,j}|}{Q_{t,j}},$$

where $r_{t,j}$ is the price return of trade j of this bond on day t, $Q_{t,j}$ is the volume of a corresponding transaction, and N_t is a total number of trades of this bond per day.[2] This definition of the Amihud measure follows the approach of Dick-Nielsen, Feldhütter, and Lando (2012) with one modification. Their approach requires at least two trades per day to compute the Amihud measure; I compute it even for days with a single trade. In this case,

[2] To see how the Amihud measure behaves on daily frequency on TRACE data relative to other illiquidity measures see Schestag et al. (2016).

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Size, mln USD	639.35	588.17	1	300	500	750	15,000
Time to maturity, years	8.84	7.89	1.00	3.26	5.88	10.12	30.43
Age, years	5.83	5.18	0.00	2.08	4.25	7.83	49.66
Duration, years	6.29	4.26	0.94	3.02	5.06	8.09	19.64
Coupon rate, pct.	5.77	1.87	0.45	4.80	5.95	7.00	15.00
Credit rating	8.36	3.17	1	6	8	10	22
Trades per bond per day	7.14	14.06	1	2	3	8	2,813
Yield to maturity, pct.	4.53	2.80	0.18	2.49	4.48	5.84	39.35
Spread, pct.	2.13	2.35	0.05	0.83	1.45	2.59	35.00
Return, pct. per day	0.02	1.27	−10.42	−0.36	0.03	0.43	9.72
Distance-to-default (DD)	0.66	0.32	0.01	0.41	0.63	0.87	8.49
Amihud measure	0.51	0.87	0.00	0.03	0.18	0.58	7.16

(a) Full-sample descriptive statistics. The age variable represents time elapsed from issuance. Duration is the Macaulay duration. Ratings are in conventional numerical score; 'AAA' corresponds to 1, 'D' corresponds to 22. For spread 5 b.p. and 35% are truncation points. The Amihud price impact measure is computed as $\frac{1}{N_t}\sum_{j=1}^{N_t}\frac{|r_{t,j}|}{Q_{t,j}}$, where $r_{t,j}$ is the price return of trade j of this bond on day t, $Q_{t,j}$ is the volume of a corresponding transaction, and N_t is a total number of trades of this bond per day. The computation of the distance-to-default variable is detailed in Appendix B.1.2.

	AAA	AA	A	BBB	BB	B	CCC	CC	C	D
Size, mln USD	550.00	800.00	500.00	499.43	400.00	400.00	350.00	275.00	175.00	360.62
Time to maturity, years	5.91	4.95	5.82	6.00	6.07	5.76	5.47	5.07	5.95	17.39
Age, years	4.35	3.33	4.04	4.18	4.93	4.95	5.51	10.82	13.36	10.96
Duration, years	5.15	4.54	5.08	5.17	5.12	4.81	4.58	4.33	4.96	10.78
Coupon rate, pct.	5.15	4.75	5.50	6.00	7.00	7.50	7.75	7.70	8.50	7.45
Trades per bond per day	4.00	5.00	3.00	3.00	4.00	4.00	5.00	6.00	4.00	5.00
Yield to maturity, pct.	3.19	3.51	3.78	4.42	6.21	7.62	9.75	13.42	13.62	14.84
Spread, pct.	0.46	0.62	0.95	1.65	3.47	4.68	6.76	10.64	11.30	10.81
Return, pct. per annum	4.46	3.62	5.02	6.89	9.41	10.87	11.56	21.66	20.26	57.14
Distance-to-default (DD)	1.14	0.94	0.77	0.58	0.43	0.37	0.25	0.18	0.21	0.09
Amihud measure	0.20	0.15	0.15	0.19	0.23	0.26	0.41	0.69	0.63	0.92
% of total	0.63	6.12	35.67	39.85	10.58	4.94	1.86	0.19	0.13	0.02
% callable	47.42	69.94	85.63	88.64	84.01	80.95	69.49	41.92	42.85	39.82

(b) Median values by credit rating except for '% of total' and '% callable'. Here, numerical ratings of Table 2.1a are aggregated to 10 letter-coded bins. Median returns here are total returns expressed in *% per annum*.

Table 2.1. Summary statistics. The full sample is 2,032,455 bond-day observations that span a period from Oct 4, 2004 to Dec 23, 2014. The sample includes only senior unsecured non-convertible fixed coupon corporate bond issues with less than 30 years to maturity. The number of unique bonds/firms in sample is 4640/775. Appendix B.1.1 details the steps of data construction. The spread in both tables is the GZ spread from Gilchrist and Zakrajšek (2012): a difference in yields to maturity between a risky bond and an imaginary risk-free bond with the exact same cash flows.

the price return is relative to a previous trade whenever it occurred.[3] Table 2.1b presents

[3] I experimented with these two definitions and found that for bond-days with at least two trades per day two definitions give very close numerical measures.

median values of the Amihud measure in the sample by credit rating. Bonds of lower credit quality tend to be less liquid in the sample.

As Table 2.1b shows, A- and BBB-rated callable bonds are predominant in the sample. The spread measure is not option-adjusted by construction; as in Gilchrist and Zakrajšek (2012), I will control for that in the EBP calculations. The median GZ spread and distance-to-default are aligned with credit ratings in an intuitive way. The higher the rating, the 'farther' the default is and the lower the spread. Ratings are also aligned (except AAA-rated and almost defaulted bonds) with median coupons, durations, and total daily returns.

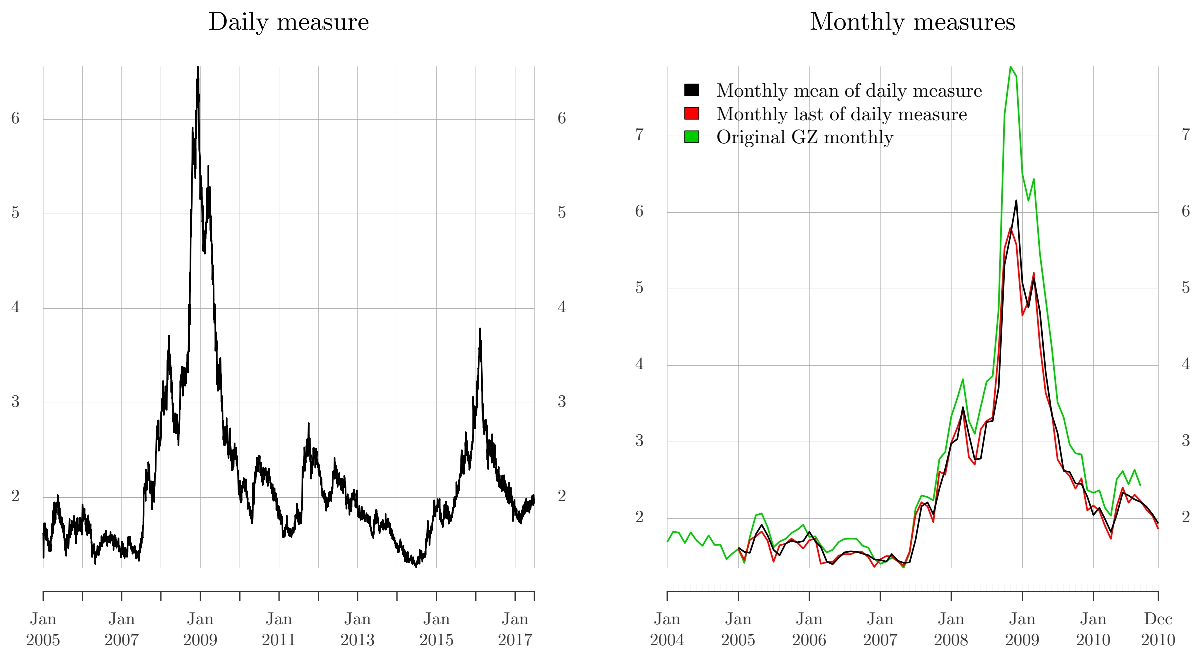


Figure 2.1. Daily and monthly measures of the aggregate GZ spread (simple cross-sectional average of the GZ spread across all bonds for each time observation). The left chart shows daily GZ spread obtained on the daily TRACE-based sample. The right chart compares it with the original monthly GZ spread from Gilchrist and Zakrajšek (2012).

My aggregate spread measure constructed on the daily data is in line with the monthly measure of Gilchrist and Zakrajšek (2012), as Figure 2.1 demonstrates. The left panel is my daily time series. For each day, the aggregate spread is a simple cross-sectional average of GZ spreads across all bonds of all firms sampled on that day. The aggregate spread is non-stationary in levels. The right panel of Figure 2.1 compares the monthly mean and last values of my daily measure with the original monthly spread from Gilchrist and Zakrajšek (2012). The three series differ a bit only during the 2008-2009 crisis; otherwise, the fit is tight. Thus, at this stage, I have obtained a larger sample with the daily data and constructed the daily GZ-spread measure, which is very close to the monthly one presented in the literature.

2.3 Measuring Excess Bond Premium

Excess bond premium (EBP) is the portion of credit spread not explained by credit risk factors. Given a panel of bonds k issued by firms i and observed at times t, and given their GZ spreads $S_{i,t}^{GZ}[k]$, bond-level $EBP_{i,t}[k]$ is computed as follows:

$$\begin{aligned}
\log\left(S_{i,t}^{GZ}[k]\right) &= \text{Factors of credit spreads} + \epsilon_{i,t}[k], \\
\hat{S}_{i,t}^{GZ}[k] &= \exp\left(\text{Part due to estimated factors} + \frac{\hat{\sigma}^2_{\epsilon_{i,t}[k]}}{2}\right), \\
EBP_{i,t}[k] &= S_{i,t}^{GZ}[k] - \hat{S}_{i,t}^{GZ}[k],
\end{aligned}$$

where $\hat{\sigma}^2_{\epsilon_{i,t}[k]}$ is the variance of residuals of the log-spread-fitting regression above. In this work, I am interested in the properties of the aggregate excess bond premium EBP_t defined for each day t as a simple cross-sectional average of $EBP_{i,t}[k]$ across all bonds of all firms.

I estimate the EBP on the daily data, unlike Gilchrist and Zakrajšek (2012) and De Santis (2017), who worked with bond–month panels. My major motivation is pronounced business cycle forecasting properties of monthly EBP established in the literature. Is it possible to extract the information about the future state of the economy beyond what we know from daily real activity measurements from credit spreads on a daily basis? Does this approach bring new information that is valuable for forecasting not only macroeconomic activity but also bond returns?

To answer these questions, I want to capture the portion of bond spreads *beyond* firm-specific credit risk, bond-specific liquidity risk, and economy-wide business risk. I directly control for bond-specific illiquidity with the daily Amihud measure and for aggregate business risk with a high-frequency *real activity* proxy. This is the daily ADS index computed and published in real time by the Philadelphia Fed.[4] The ADS index based on Aruoba, Diebold, and Scotti (2009) is a smoothed business cycle state derived from a mixed-frequency state-space linear model for six real-valued variables: initial jobless claims, payroll employment, industrial production, personal income less transfer payments, manufacturing and trade sales, and gross domestic product. The ADS index contains neither bond nor stock market data as inputs.

I benchmark my EBP measure on Gilchrist and Zakrajšek (2012). Their EBP is correlated with economy-wide business risk, and bond-specific illiquidity is controlled only with monthly bond characteristics. De Santis (2017) controlled for aggregate business risk when measuring EBP, but he estimated it on monthly European multi-country data. This summarizes the differences in my preferred spread-fitting model relative to the Gilchrist and Zakrajšek (2012) model:

Original GZ models:

$$\log\left(S_{it}^{GZ}[k]\right) = \beta \cdot DD_{it} + \text{(Proxies for recovery rate and liquidity)} + \text{(Call adjustment)} + \\ + \text{(Industry and rating FE)} + \epsilon_{it}[k].$$

My preferred models:

$$\log\left(S_{it}^{GZ}[k]\right) = \beta \cdot DD_{it} + \text{(Proxies for recovery rate and liquidity)} + \text{(Call adjustment)} + \\ + \underbrace{\gamma \cdot ADS_t + \eta \cdot AMH_{it}[k]}_{\text{Daily business cycle and liquidity}} + \text{(Industry and rating FE)} + \epsilon_{it}[k],$$

[4]For details on the ADS index see Appendix B.2 and `https://goo.gl/mZJ5Sj`. Many alternative daily aggregate business risk proxies exist, for instance, the Economic Policy Uncertainty index of Baker, Bloom, and Davis (2016). I opted for the ADS mainly because of a long history of ADS vintages readily available at the Philadelphia Fed web page. These historical vintages allow me to perform out-of-sample analysis in Sections 2.5.2 and 2.5.4.

where DD_{it} is the distance-to-default of firm i at time t (proxy for idiosyncratic credit risk), ADS_t is the aggregate business activity index at day t, and $AMH_{it}[k]$ is the Amihud measure. In the following sections, I apply both approaches on the bond-day panel and investigate the differences in the resulting EBPs.

Table 2.2 presents the estimated models over the entire sample. Model 1 is the basic model with corporate credit risk factors on the right-hand side. In Models 2 and 3, I consecutively add aggregate business activity and liquidity factors. Models 1 to 3 have the simplest possible call option adjustment: a constant that is identical for all bonds at all times. Models 4 to 6 introduce interactions of call dummy with yield curve factors and bond characteristics to possibly better capture the time variation in the issuer's desire to call an issue before maturity. Model 4 is the benchmark Gilchrist and Zakrajšek (2012) model. I see Model 6, which extends Model 4 with daily aggregate business activity and liquidity factors, as the alternative model.

All models in Table 2.2 have high explanatory power for log spreads. Even the simplest model, Model 1, explains around 72% of the log spreads variation in the data. More elaborate call option adjustment (Model 4) increases this share by 2.5 percentage points. Aggregate business activity and liquidity factors (Model 6) add another 4.5 percentage points to the share of explained log spreads variation, which reaches 79%.

As Table 2.2 shows, aggregate business risk and bond illiquidity are significant predictors of credit spreads. Coefficients on the ADS business cycle index and the Amihud measure are statistically significant across all specifications. They do not vary much from one model to another and have reasonable signs. Business cycle upturns are associated with lower credit spreads, and more illiquid bonds have higher spreads.

Most interaction variables of the call dummy with yield curve factors and bond characteristics introduced in Models 4–6 for the purpose of call option adjustment are statistically significant, as Table 2.2 shows. Observe for Models 5 and 6 that when the yield curve moves up and becomes steeper, the spreads tend to become lower. This finding can be explained as

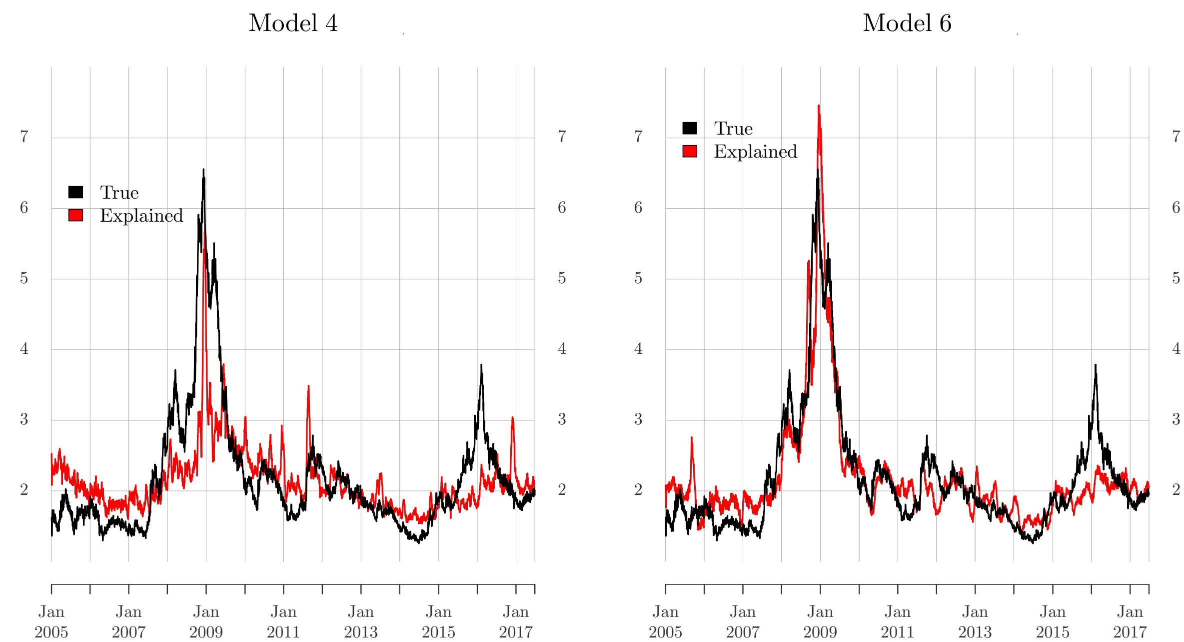
	Dependent variable: $\log(Spread_{it}[k])$					
	(1)	(2)	(3)	(4)	(5)	(6)
$-DD_{it}$	0.846***	0.631***	0.625***	0.680***	0.487***	0.478***
	(0.037)	(0.036)	(0.035)	(0.061)	(0.059)	(0.058)
$\log(DUR_{it}[k])$	0.311***	0.322***	0.311***	0.304***	0.313***	0.297***
	(0.011)	(0.011)	(0.011)	(0.022)	(0.021)	(0.020)
$\log(PAR_{it}[k])$	−0.078***	−0.076***	−0.070***	−0.050***	−0.053***	−0.046***
	(0.012)	(0.011)	(0.011)	(0.016)	(0.017)	(0.017)
$\log(CPN_i[k])$	0.515***	0.451***	0.449***	0.620***	0.550***	0.544***
	(0.020)	(0.019)	(0.019)	(0.063)	(0.061)	(0.060)
$\log(AGE_{it}[k])$	−0.004	0.013**	0.008	0.004	0.031	0.029
	(0.006)	(0.006)	(0.006)	(0.028)	(0.027)	(0.027)
$CALL_i[k]$	0.036	0.046**	0.051**	0.611***	0.765***	0.784***
	(0.023)	(0.022)	(0.021)	(0.233)	(0.244)	(0.240)
ADS_t		−0.266***	−0.261***		−0.255***	−0.251***
		(0.007)	(0.007)		(0.007)	(0.007)
$AMH_{it}[k]$			0.051***			0.053***
			(0.002)			(0.002)
$-DD_{it} \cdot CALL_i[k]$				0.106*	0.113*	0.116*
				(0.064)	(0.061)	(0.060)
$\log(DUR_{it}[k]) \cdot CALL_i[k]$				0.015	0.009	0.014
				(0.019)	(0.018)	(0.018)
$\log(PAR_{it}[k]) \cdot CALL_i[k]$				−0.032*	−0.032*	−0.033*
				(0.017)	(0.019)	(0.018)
$\log(CPN_i[k]) \cdot CALL_i[k]$				−0.144**	−0.087	−0.081
				(0.065)	(0.063)	(0.062)
$\log(AGE_{it}[k]) \cdot CALL_i[k]$				−0.001	−0.025	−0.028
				(0.027)	(0.026)	(0.026)
$LEV_t \cdot CALL_i[k]$				−0.012	−0.034***	−0.037***
				(0.009)	(0.008)	(0.008)
$SLP_t \cdot CALL_i[k]$				−0.00003	−0.025**	−0.026**
				(0.012)	(0.010)	(0.010)
$CRV_t \cdot CALL_i[k]$				0.030**	−0.035***	−0.035***
				(0.014)	(0.012)	(0.012)
$VOL_t \cdot CALL_i[k]$				2.052***	0.720***	0.703***
				(0.094)	(0.055)	(0.054)
Industry FE	YES	YES	YES	YES	YES	YES
Credit rating FE	YES	YES	YES	YES	YES	YES
Observations	2,756,326	2,756,326	2,756,326	2,756,326	2,756,326	2,756,326
Adjusted R^2	0.751	0.797	0.800	0.767	0.801	0.804
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table 2.2. Candidate explanatory models for the bond k of firm i – day t **panel of credit spreads** for the entire sample (Oct 4, 2004 – Dec 23, 2014). The dependent variable is the log of GZ spread. DD is the distance-to-default, DUR is duration, PAR is amount outstanding, CPN is the coupon rate, AGE is time elapsed from issuance, and $CALL$ is a callable bond dummy. ADS is the Aruoba-Diebold and Scotti aggregate activity index, AMH is the Amihud liquidity measure. LEV, SLP, and CRV are correspondingly level, slope, and curvature yield curve factors, and VOL is the realized volatility of the 10-year rate (30-day moving average). See Appendix B.2 for the details on explanatory variables. All models include industry (the first two digits of the NAICS code) and credit rating (22-grade numeric scale) fixed effects. Standard errors are clustered in both firm i and time t dimensions. Model (4) is a benchmark model (Gilchrist and Zakrajšek, 2012), Model (6) is used as an alternative model throughout the rest of the chapter.

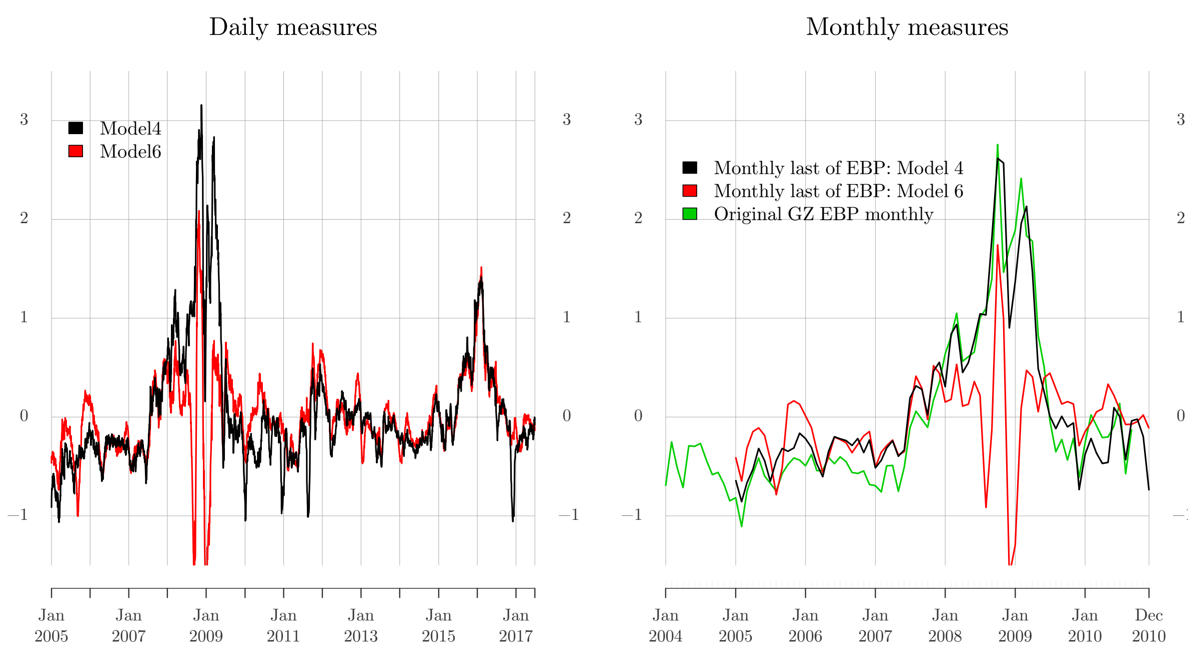
follows: the probability of an early call decreases when rates become higher (fewer incentives for an issuer to refinance at higher rates); hence, an early call premium drops and callable bonds tend to become more expensive. The importance and significance of call option adjustment make Gilchrist and Zakrajšek (2012) argue that Model 4 is superior to Model 1 on their data, I get the same result on my data. My primary interest, however, is in comparison of Models 4 and 6, which I turn to now.

Figure 2.2a compares the goodness of fit of Models 4 (benchmark Gilchrist and Zakrajšek, 2012) and 6 (my preferred model) to actual aggregate spreads. Model 6 captures time series variation in daily spreads much better than Model 4, especially in years 2008 and 2009, and this is due to only two additional factors: the state of the business cycle and bond liquidity. The left panel of Figure 2.2b presents the same result in terms of the EBP. An unexplained increase in credit spread during the subprime crisis is significantly smaller and shorter in time according to my preferred model; hence, the state of the business cycle is a factor of aggregate credit spread even on a daily frequency. The right panel of Figure 2.2b compares monthly EBP values of my preferred daily EBP (Model 6) with the original monthly EBP series from Gilchrist and Zakrajšek (2012) and confirms this finding.

The significance of business cycle and liquidity as factors of credit spreads survives the truncation of the data sample. Table 2.3 compares performance of Models 4 and 6 in subsamples of either investment-grade or high-yield bonds. The models explain spreads of investment-grade bonds much better than high-yield ones. Spreads of the riskiest bonds are probably non-linear in the distance-to-default: coefficients on the DD variable in columns 3 and 4 (high-yield bonds) of Table 2.3 are roughly twice the corresponding coefficients in columns 1 and 2 (investment-grade bonds). Yet, business cycle and liquidity factors are still significant for both types of bonds, and coefficients on these variables are not much different from the full-sample specifications. More importantly, business cycle and liquidity factors survive complete deletion of observations between Jan 2008 and Dec 2008 (inclusive) from the sample. Columns 5 and 6 of Table 2.3 present these estimations. Both coefficients do not change much relative to full-sample specifications and improve the explanatory power of Model 6 as measured by the R^2 compared to Model 4.



(a) Daily time series of true and fitted GZ spread from by Model 4 on the left (Gilchrist and Zakrajšek, 2012) and Model 6 on the right (my preferred model).



(b) Monthly time series of EBP from alternative models.

Figure 2.2. Fitted spread and EBP (the residual portion of spread) that are computed with different models of Table 2.2 in comparison with the original Gilchrist and Zakrajšek (2012) EBP measure.

	Dependent variable: log $\left(S_{it}^{GZ}[k]\right)$					
	M4	M6	M4	M6	M4	M6
	IG bonds: all years		HY bonds: all years		All bonds, ex. year 2008	
$-DD_{it}$	0.682***	0.464***	1.127***	0.968***	0.594***	0.459***
	(0.060)	(0.054)	(0.176)	(0.164)	(0.063)	(0.060)
$\log(DUR_{it}[k])$	0.325***	0.323***	0.208***	0.196***	0.332***	0.322***
	(0.022)	(0.020)	(0.030)	(0.030)	(0.024)	(0.023)
$\log(PAR_{it}[k])$	−0.076***	−0.071***	0.012	0.014	−0.053***	−0.048***
	(0.018)	(0.019)	(0.022)	(0.022)	(0.017)	(0.017)
$\log(CPN_i[k])$	0.636***	0.554***	0.402**	0.334**	0.547***	0.509***
	(0.057)	(0.055)	(0.158)	(0.153)	(0.063)	(0.061)
$\log(AGE_{it}[k])$	−0.004	0.023	−0.034	−0.007	0.033	0.043
	(0.029)	(0.028)	(0.045)	(0.039)	(0.029)	(0.029)
$CALL_i[k]$	0.302	0.518*	0.363	0.510	0.759***	0.856***
	(0.255)	(0.270)	(0.445)	(0.432)	(0.242)	(0.244)
ADS_t		−0.255***		−0.210***		−0.217***
		(0.007)		(0.012)		(0.007)
$AMH_{it}[k]$		0.048***		0.053***		0.055***
		(0.002)		(0.005)		(0.002)
$-DD_{it} \cdot CALL_i[k]$	0.097	0.113*	−0.198	−0.158	0.126*	0.125**
	(0.066)	(0.060)	(0.161)	(0.153)	(0.065)	(0.062)
$\log(DUR_{it}[k]) \cdot CALL_i[k]$	0.019	0.013	−0.015	−0.007	0.003	0.005
	(0.020)	(0.018)	(0.029)	(0.029)	(0.021)	(0.019)
$\log(PAR_{it}[k]) \cdot CALL_i[k]$	−0.008	−0.012	−0.041	−0.043	−0.033*	−0.034*
	(0.019)	(0.021)	(0.027)	(0.027)	(0.018)	(0.019)
$\log(CPN_i[k]) \cdot CALL_i[k]$	−0.192***	−0.122**	0.027	0.098	−0.084	−0.052
	(0.060)	(0.059)	(0.162)	(0.156)	(0.065)	(0.064)
$\log(AGE_{it}[k]) \cdot CALL_i[k]$	0.011	−0.017	0.033	−0.004	−0.030	−0.044
	(0.029)	(0.027)	(0.045)	(0.039)	(0.028)	(0.028)
$LEV_t \cdot CALL_i[k]$	0.011	−0.017**	−0.071***	−0.085***	−0.046***	−0.052***
	(0.009)	(0.008)	(0.015)	(0.015)	(0.008)	(0.008)
$SLP_t \cdot CALL_i[k]$	−0.015	−0.041***	0.034**	0.014	−0.019	−0.026**
	(0.013)	(0.011)	(0.015)	(0.014)	(0.012)	(0.011)
$CRV_t \cdot CALL_i[k]$	0.034**	−0.039***	0.001	−0.032	0.014	−0.026**
	(0.015)	(0.012)	(0.024)	(0.021)	(0.014)	(0.013)
$VOL_t \cdot CALL_i[k]$	2.094***	0.715***	1.878***	0.808***	1.899***	0.763***
	(0.097)	(0.054)	(0.119)	(0.099)	(0.092)	(0.052)
Industry FE	YES	YES	YES	YES	YES	YES
Credit rating FE	YES	YES	YES	YES	YES	YES
Observations	2,267,713	2,267,713	488,613	488,613	2,600,016	2,600,016
Adjusted R^2	0.699	0.751	0.543	0.592	0.784	0.806
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table 2.3. Model 4 (M4 columns) and Model 6 (M6 columns) from Table 2.2 recomputed over **different sub-samples of the entire sample**. The first two columns are only investment-grade bonds over the entire sample, the second two columns are high-yield bonds over the entire sample, and the last two columns are all bonds but excluding all days in year 2008. Dependent variable is the log of GZ spread. Explanatory variables are as in Table 2.2. Standard errors are clustered in both firm i and time t dimensions.

An alternative way to establish the link between aggregate business risk and the portion of spreads beyond corporate bond credit risk is presented in Appendix B.3. There I first introduce, following d'Avernas (2017), time fixed effect in log spread fitting models (and remove the ADS). Then I project this estimated time fixed effect on the ADS in a univariate time-series regression to demonstrate that the latter explains significantly around 63% of the variation of the former on the daily frequency.

In this section, I have demonstrated that aggregate business risk and bond liquidity risk are significant factors of credit spreads in addition to corporate credit risk. Does the residual spread that is free from all these sources of risk (EBP of Model 6) still forecast macro as the benchmark EBP measure (Model 4)? Section 2.4 answers this question.

2.4 Forecasting the Business Cycle

I explore the forecasting properties of the EBP with respect to business activity by running predictive models for monthly industrial production, payroll employment, and the unemployment rate similar to the ones in Gilchrist and Zakrajšek (2012). Here, I use month-end values of my daily EBP measures obtained in Section 2.3. The regressions are:

$$\nabla^h Y_{t+h} = \alpha + \sum_{i=1}^{p} \beta_i \nabla Y_{t-i} + \underbrace{\gamma_1 RFF_t + \gamma_2 TS_t}_{\text{Real Fed funds rate and term spread}} + \underbrace{\gamma_3 S_t^{GZ}}_{\text{True GZ spread}} + \epsilon_{t+h},$$

$$\nabla^h Y_{t+h} = \alpha + \sum_{i=1}^{p} \beta_i \nabla Y_{t-i} + \gamma_1 RFF_t + \gamma_2 TS_t + \underbrace{\gamma_3 \hat{S}_t^{GZ} + \gamma_4 EBP_t}_{\text{Fitted GZ spread and EBP}} + \epsilon_{t+h},$$

where $\nabla^h Y_{t+h}$ is either $\log Y_{t+h} - \log Y_{t-1}$ the growth rate of industrial production/payroll employment or $Y_{t+h} - Y_{t-1}$ the change in unemployment rate. The right-hand side variables (apart from a constant and the dependent variable lags) capture different components of the real cost of borrowing through the corporate bond market for an average U.S. bond-issuing firm.[5] The literature has established long ago that, in such models, credit spreads are significant predictors for different left-hand side indicators and forecasting horizons. Gilchrist and Zakrajšek (2012) demonstrated that the predictive power of spreads is rather due to the residual spread than the fitted spread component. I revisit this result with my preferred measure of the EBP.

[5]For details on the right-hand side variables see Appendix B.2.

	Industrial production			Unemployment rate			Payroll employment		
	–	M4	M6	–	M4	M6	–	M4	M6
Real Fed funds rate	0.32	0.28	0.27	0.06	0.06	0.05	0.01	0.01	0.004
	(0.28)	(0.23)	(0.30)	(0.05)	(0.05)	(0.03)	(0.06)	(0.06)	(0.05)
Term spread	−0.53	−0.42	−0.40	0.03	0.04	0.03	−0.07	−0.07	−0.06
	(0.38)	(0.34)	(0.35)	(0.05)	(0.05)	(0.03)	(0.07)	(0.07)	(0.05)
GZ spread	−1.62***			0.52***			−0.41***		
	(0.48)			(0.08)			(0.15)		
Fitted GZ		−0.82*	−2.07***		0.55***	0.60***		−0.41***	−0.58***
		(0.43)	(0.46)		(0.08)	(0.04)		(0.13)	(0.08)
EBP		−2.27***	−0.97**		0.51***	0.26**		−0.41***	−0.19
		(0.64)	(0.45)		(0.11)	(0.10)		(0.16)	(0.13)
Adjusted R^2	0.60	0.63	0.63	0.71	0.70	0.77	0.84	0.84	0.89
Note:	*p<0.1; **p<0.05; ***p<0.01								

(a) 3 months ahead

	Industrial production			Unemployment rate			Payroll employment		
	–	M4	M6	–	M4	M6	–	M4	M6
Real Fed funds rate	0.62	0.53	0.95	0.04	−0.02	0.03	0.07	0.08	0.06
	(0.52)	(0.40)	(0.60)	(0.10)	(0.12)	(0.09)	(0.15)	(0.15)	(0.12)
Term spread	−0.90	−0.68	−1.36**	0.14	0.19*	0.14*	−0.20	−0.20	−0.19
	(0.56)	(0.44)	(0.59)	(0.09)	(0.10)	(0.08)	(0.17)	(0.16)	(0.14)
GZ spread	−2.44***			0.80***			−0.68***		
	(0.76)			(0.14)			(0.24)		
Fitted GZ		−0.84	−2.63**		0.45*	0.89***		−0.49	−0.87***
		(0.73)	(1.23)		(0.24)	(0.14)		(0.32)	(0.16)
EBP		−3.72***	−1.38*		0.76***	0.51***		−0.73**	−0.39
		(1.13)	(0.80)		(0.18)	(0.16)		(0.30)	(0.25)
Adjusted R^2	0.47	0.51	0.46	0.70	0.72	0.73	0.79	0.79	0.81
Note:	*p<0.1; **p<0.05; ***p<0.01								

(b) 6 months ahead

	Industrial production			Unemployment rate			Payroll employment		
	–	M4	M6	–	M4	M6	–	M4	M6
Real Fed funds rate	0.57	0.38	0.55	0.03	0.04	0.03	0.14	0.15	0.08
	(0.58)	(0.43)	(0.54)	(0.17)	(0.15)	(0.15)	(0.26)	(0.23)	(0.21)
Term spread	−1.51	−1.06	−1.49*	0.40**	0.37***	0.40**	−0.55*	−0.53*	−0.43*
	(0.93)	(0.66)	(0.87)	(0.18)	(0.14)	(0.16)	(0.28)	(0.27)	(0.26)
GZ spread	−3.89			1.18***			−1.15**		
	(2.72)			(0.31)			(0.49)		
Fitted GZ		−0.67	−3.71		0.88***	1.26***		−0.68	−1.44***
		(1.75)	(2.61)		(0.16)	(0.36)		(0.59)	(0.42)
EBP		−6.46*	−4.27		1.32***	0.91**		−1.27**	−0.92
		(3.55)	(2.84)		(0.51)	(0.37)		(0.58)	(0.57)
Adjusted R^2	0.26	0.33	0.26	0.54	0.54	0.54	0.62	0.62	0.62
Note:	*p<0.1; **p<0.05; ***p<0.01								

(c) 12 months ahead

Table 2.4. Forecasting regressions for the log growth rate of **industrial production**, change in the **unemployment rate**, and the log growth rate of **payroll employment** on different horizons (*not* annualized) with either true spread or fitted spread and the EBP (excess bond premium) as explanatory variables. The EBP is from two alternative models of Table 2.2: Models 4 and 6 (columns 'M4' and 'M6' correspondingly). Real Federal funds rate is the difference between nominal rate and realized 12-month inflation (one month prior to a rate observation), Term spread is the difference between 3-month and 10-year Treasury zero coupon rates. See Appendix B.2 for the details on explanatory variables. Each regression also has a constant and an automatically selected number of lags (based on the AIC) of the dependent variable (also not reported). Sample period is monthly from Oct 2004 to Dec 2014. Standard errors are Newey and West (1987) HAC estimates.

Table 2.4 demonstrates that high spreads today are indeed associated with lower future industrial production and higher future unemployment in my sample.[6] The columns titled '–' estimate models with GZ spread as an explanatory variable without splitting it into explained and unexplained parts. For the industrial production, the unemployment rate and the payroll employment on all horizons (except for one-year ahead industrial production), the spread is indeed a strong predictor of future macroeconomic activity with reasonable signs.[7]

As 'M4' columns of Table 2.4 show, the EBP computed as in Gilchrist and Zakrajšek (2012) is indeed a stronger predictor of future macro activity than the explained portion of spread ('fitted spread'). For the industrial production, the EBP of Model 4 is a significant predictor, and the fitted spread is not. Speaking about economic significance, the absolute value of the coefficients on EBP is 4-7 times higher depending on the forecasting horizon. For the employment-related variables, both the EBP and the fitted spread are statistically significant predictors, but the economic significance of changes in EBP for future employment trends is, again, substantially higher than of changes in fitted spreads, especially on longer horizons.

The predictive power of the EBP becomes considerably lower once I switch to residual spreads free from corporate default risk, aggregate business risk, and bond liquidity risk. This result is the most pronounced for the industrial production. Observe in the column titled 'M6' of Table 2.4a that for three-month ahead growth of the industrial production, the fitted spread is now a significant predictor, and the EBP is not. Compared to 'M4' column, not only the significance but also the magnitude of coefficients on the fitted spread and the EBP has changed considerably. The same result applies to the 6-month ahead industrial production, Table 2.4b shows. At the 12-month horizon, Table 2.4c, neither of the two

[6] This table echoes Table 6 of Gilchrist and Zakrajšek (2012).

[7] I use Newey and West (1987) heteroskedasticity and autocorrelation consistent standard errors in forecasting regressions with overlapping observations in Sections 2.4 and 2.5 of the chapter. I also ran all the estimations with Hodrick (1992) standard errors instead and found that in my sample Newey-West standard errors are bigger than Hodrick's ones in vast majority of cases. Hence I reject the 'no predictability' null less frequently using the Newey-West errors.

components of the spread is a significant predictor of industrial production. Table 2.4 also presents similar results for the unemployment rate and the payroll employment. Here, in 'M6' columns, both components of the spread are still statistically significant predictors of employment trends, but the economic significance of the fitted spread is now much higher than of the EBP (especially on the 3-month horizon, where coefficients on the fitted spread are roughly twice higher in absolute value than the coefficients on the EBP). Hence, switching from Model 4 to Model 6 increases both statistical and economic significance of the fitted spread and shrinks the significance of the EBP in forecasting economic activity.[8]

The results discussed in this section so far go through if one compares instead predictive models with the fitted spread and the EBP of Models 1 and 3 of Table 2.2 (not reported). This case refers to the EBP estimations when call option adjustment is just the loading on the call dummy, same for all callable bond at all times. Hence, the reduction in predictive power of the EBP for future macroeconomic activity is not due to the chosen method of embedded call option adjustment.

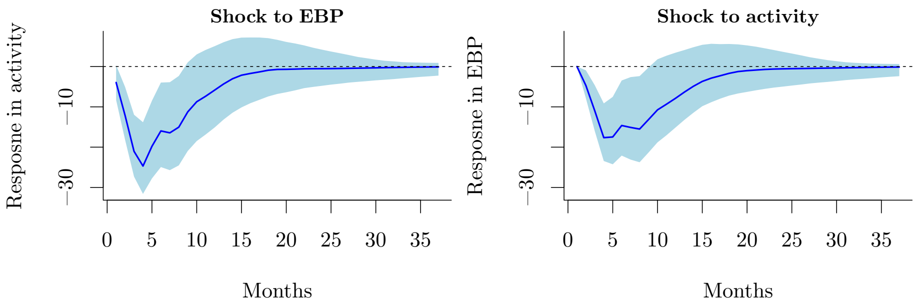
I interpret the findings of this section as follows: the daily EBP measure contains information that is relevant for predicting the future state of the economy, but this information can be also derived from a readily available daily business activity measure, such as the ADS index. The forecasting power of the residual spread that is free from the corporate default risk only (EBP of Model 4 or Model 1) is mostly due to the persistence of the business cycle itself.[9] Once one further projects spreads on the aggregate activity and illiquidity measures, the forecasting power of the residual component (EBP of Model 6 or Model 3) goes away completely for some macro indicators and falls considerably for the others.[10]

---

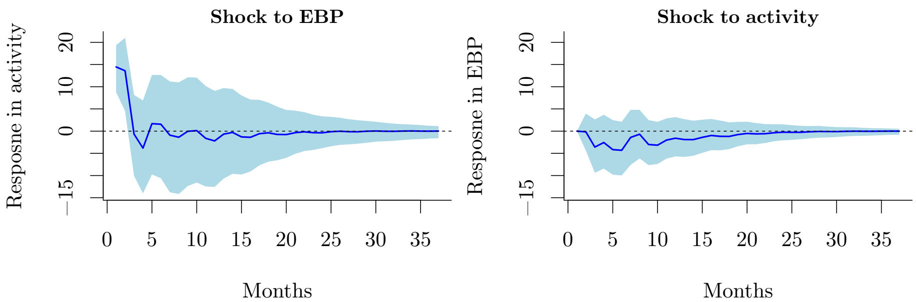
[8]This result is reinforced by the lower variability of the EBP estimated by Model 6 compared to Model 4: large deviations of the EBP of Model 6 from its mean are less probable per se.

[9]For estimations of the U.S. business cycle persistence see Mariano and Murasawa (2003).

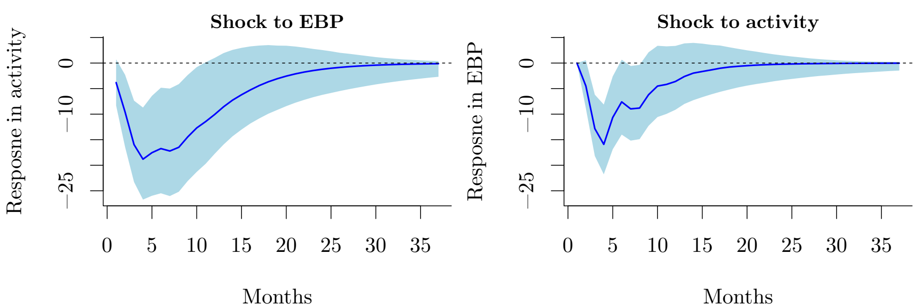
[10]Gilchrist and Zakrajšek (2012) also considered structural shocks to their EBP measure in a quarterly eight-variable macro SVAR model and interpreted the shocks as 'EBP shocks orthogonal to the business cycle'. However, this interpretation hinges on the identification of the SVAR model by exclusion restrictions. Their identification yields significant effects of 'EBP shocks orthogonal to the business cycle' on activity. I believe that it is better to directly control for the state of the business cycle at the stage of the EBP estimation. This approach leads to a different conclusion regarding the forecasting power of the EBP relative to the fitted spread.



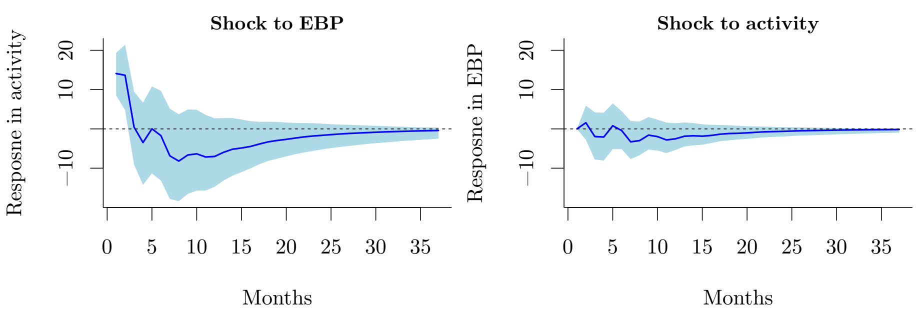
(a) Here EBP is computed with **Model 1** of Table 2.2.



(b) Here EBP is computed with **Model 3** of Table 2.2.



(c) Here EBP is computed with **Model 4** of Table 2.2.



(d) Here EBP is computed with **Model 6** of Table 2.2.

Figure 2.3. Orthogonalized impulse-response functions (IRFs) to one standard deviation shocks from bi-variate monthly **VAR models of business activity (the ADS index) and the EBP**. Monthly ADS and EBP are the latest daily observations per month. The models include a number of lags selected by AIC (required to be less or equal to 12) and a constant. The EBP is based on Models 1, 3, 4, and 6 of Table 2.2. Shaded areas are 95% bootstrapped confidence bands (10'000 runs). Sample period is from Oct 2004 to Dec 2014.

Bi-variate monthly vector autoregression (VAR) models on the EBP and the ADS activity index provide supporting evidence for such an interpretation.[11] I estimate these VARs to capture possible time series interdependence of activity and the EBP. Figure 2.3 presents orthogonalized impulse response functions from the estimated models, with the EBPs of Models 1, 3, 4, and 6 on Figures 2.3a, 2.3b, 2.3c, and 2.3d correspondingly. The response of activity on the EBP shock on the left panel of Figure 2.3a (the EBP of Model 1) shows that unexpected jumps in the EBP today imply significantly lower business activity up to nine months ahead, and vice versa. The EBP-to-activity pass-through remains the same when I consider the EBP from Model 4 instead, Figure 2.3c shows, hence this finding is not due to the chosen method of call option adjustment. However, once I consider the EBP free from liquidity risk and aggregate business risk (Models 3 and 6), the link between the EBP and activity breaks up. Figures 2.3b and 2.3d show that now shocks to the EBP do not affect activity significantly over horizons longer than several months (and over these shorter horizons the effect has a counter-intuitive sign). There is no significant effect in the opposite direction either. These results corroborate the findings of this chapter: the portion of credit spreads explained by firm-specific credit risk, economy-wide business risk and bond-specific liquidity risk does a good job in forecasting the future macroeconomic state, and the residual portion of spreads is less important for macro forecasting.

2.5 Forecasting Corporate Bond Returns

In this section, I investigate two questions: whether the EBP contains any information relevant for forecasting bond returns or not, and how the EBP relates to other known bond pricing factors. The motivation for this part comes from the decomposition of credit spreads by Nozawa (2017). This paper shows that the Campbell-Shiller decomposition applied to

[11] On monthly frequency, both the ADS and the EBP are stationary time series over the years 2004–2014. I obtain monthly values of these series by taking the latest daily observation per month. Taking monthly means instead doesn't change the results.

corporate bond spreads (under mild assumptions about losses in default) yields:

$$S_t = \underbrace{\mathbb{E}_t\left[\sum_{i=1}^{\infty}\rho^{i-1}r^e_{t+i}\right]}_{\text{Risk premium}} + \underbrace{\mathbb{E}_t\left[\sum_{i=1}^{\infty}\rho^{i-1}l_{t+i}\right]}_{\text{Expected credit loss}} + \text{Const},$$

where ρ is the steady-state price-coupon ratio, r^e is excess bond return and l is credit loss. I want to think of the empirical decomposition of GZ spreads into an explained part and the EBP as one particular model-based method to reinterpret the Campbell-Shiller decomposition above. The EBP is interpreted in this case as the credit risk premium (i.e., conditional expectation of future excess corporate bond returns). Then, it is natural to ask whether the EBP forecasts actual future returns.

2.5.1 EBP and Excess Corporate Bond Market Returns

Building on regression models of Section 2.4, I estimate the following forecasting models on the daily data:

$$R_{t:t+h} = \alpha + \beta R_{t-h:t} + \underbrace{\gamma_1 LVL_t + \gamma_2 SLP_t + \gamma_3 CRV_t}_{\text{level, slope and curvature factors}} + \underbrace{\gamma_4 S^{GZ}_t}_{\text{true GZ spread}} + \epsilon_{t+h},$$

$$R_{t:t+h} = \alpha + \beta R_{t-h:t} + \gamma_1 LVL_t + \gamma_2 SLP_t + \gamma_3 CRV_t + \underbrace{\gamma_4 \hat{S}^{GZ}_t + \gamma_5 EBP_t}_{\text{Fitted GZ and EBP}} + \epsilon_{t+h},$$

where $R_{t:t+h} = \sum_{i=1}^{h} R_{t+i}$ are cumulative excess log returns on a diversified bond portfolio h days ahead.[12] I consider the range of horizons from 1 day to 90 days to ensure the stationarity of the returns series on the left-hand side.[13] The left-hand side returns are for one of the two alternative bond market portfolios: the value-weighted portfolio of in-sample TRACE

[12] Returns are total returns here, they account for both price changes and accrued interest.

[13] Cumulative returns are non-stationary on horizons beyond roughly 90 days. Hence one needs to test for cointegration between returns and potential predictors on these longer horizons instead. I did that, and the tests didn't reject the null (no cointegration, i.e. no predictability for returns coming from the fitted spread, the EBP, or the yield curve factors).

bonds and the portfolio of investment-grade bonds in the Barclays Aggregate U.S. corporate bond index.[14]

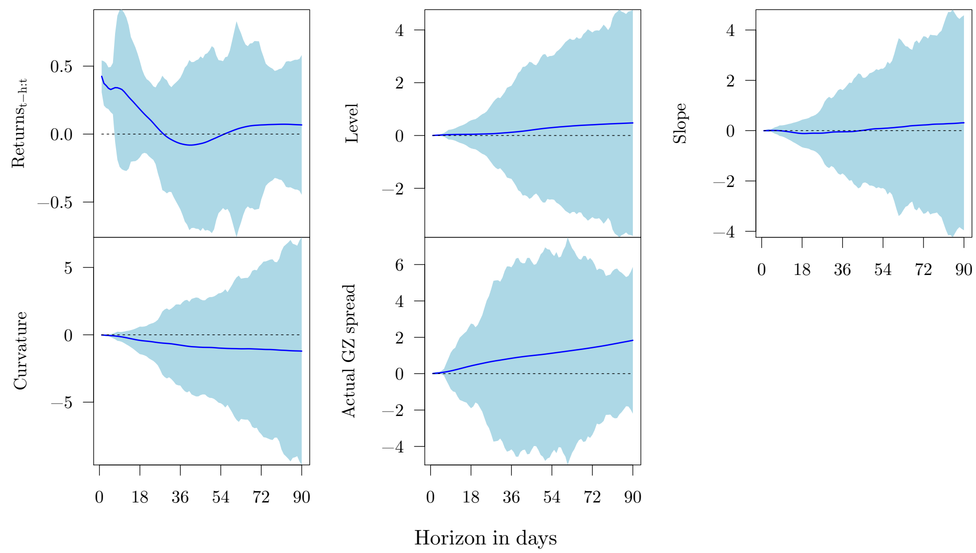
My findings are as follows: the actual GZ spread is not a significant predictor of bond market returns. Figure 2.4a presents the estimates of the various parameters in the model with actual GZ spread on the right and their significance over different horizons. None of the factors significantly predicts cumulative returns in such a model on horizons up to 90 days.

The residual spread free from corporate default risk only (EBP of Model 4) is not a predictor of bond market returns either. Figure 2.4b presents these estimations, and here, again, none of the factors is significant at horizons below 70 business days. For 70–90 days ahead the fitted spread is a significant in-sample predictor of cumulative returns, but, as I show later, this result is not robust to alternative specifications of market returns and the EBP. The bottom line of the estimations presented in Figure 2.4 is as follows: if there is any information in aggregate spreads relevant for forecasting future excess bond market returns at all, it can hardly be extracted using the EBP correlated with the state of the business cycle.

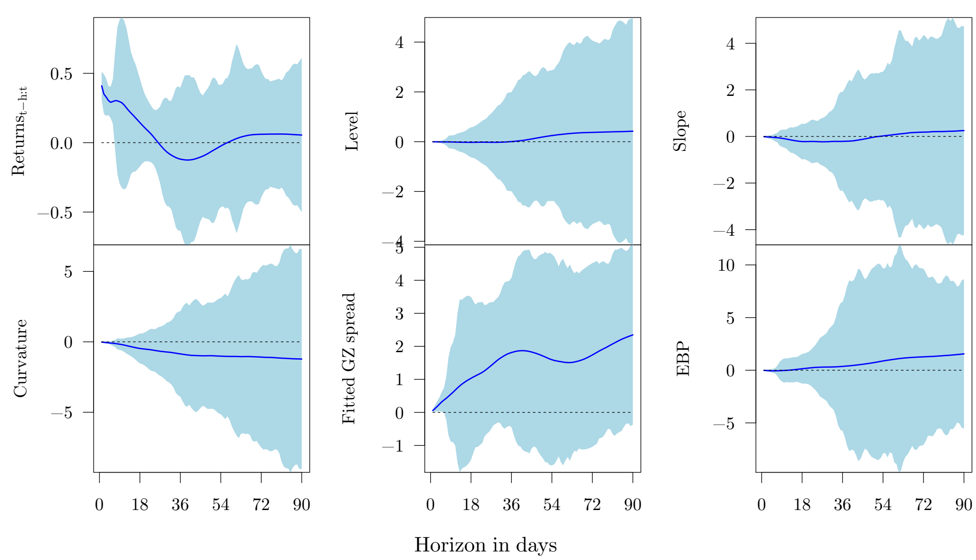
In contrast, once one switches to the residual spread free from aggregate business risk and bond liquidity risk (EBP of Model 6), such bond premiums, unlike the EBP of Gilchrist and Zakrajšek (2012), turn out to be a significant predictor of bond market returns. Figure 2.5a presents the results of such forecasting models. For all horizons between 40 and 60 days ahead, the EBP and only the EBP is a significant predictor of excess bond market returns. Economic significance of the EBP for future returns is high as well. A 10 basis points (b.p.) rise in the EBP today implies almost 40 b.p. of excess bond market return over the next two-three month. To give a sense of scale, the average absolute daily change in the EBP in my sample is 4 b.p. with a standard deviation of 5 b.p. The adjusted R^2 of return forecasting regression at the 50-day horizon is 0.52.

[14] All subsequent results were also obtained fot the equally-weighted portfolio of TRACE bonds (not presented here).

Parameter estimates for cumulative returns on different horizons



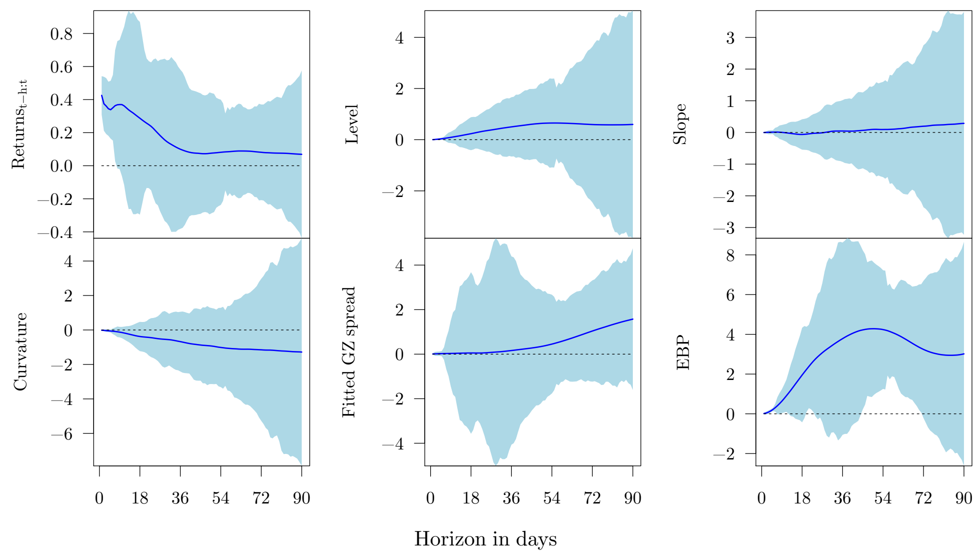
(a) Dependent variable: returns on **TRACE portfolio** of bonds; actual GZ spread as one of explanatory variables.



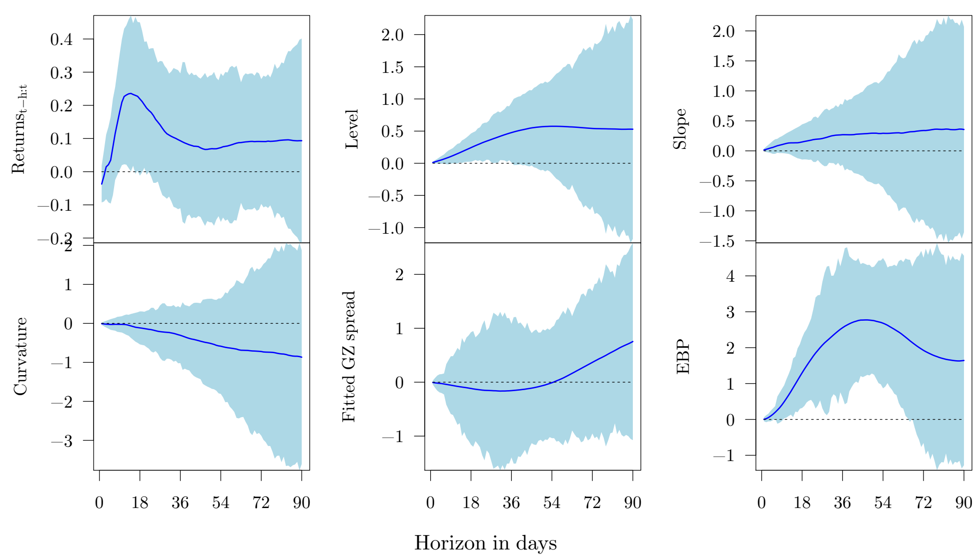
(b) Dependent variable: returns on **TRACE portfolio** of bonds; fitted GZ spread and EBP of **Model 4** as explanatory variables.

Figure 2.4. Estimated **forecasting regressions for cumulative bond market excess returns**. Forecasting horizons are on horizontal axes. Market returns are log returns (*not* annualized) on the value-weighted portfolio of TRACE bonds. Explanatory variables are on vertical axes. See Appendix B.2 for the details on explanatory variables. Each point on a solid line on each chart is the OLS-estimate from a corresponding regression. Shaded areas around are two standard errors of the estimates. The standard errors are heteroskedasticity and autocorrelation consistent estimates of Newey and West (1987). Each model also includes a constant (not reported). The sample is daily from Oct 4, 2004 to Dec 23, 2014.

Parameter estimates for cumulative returns on different horizons



(a) Dependent variable: returns on **TRACE portfolio** of bonds; fitted GZ spread and EBP of **Model 6** as explanatory variables.



(b) Dependent variable: returns on **Barclays Aggregate U.S. corporate bond index**; fitted GZ spread and EBP of **Model 6** as explanatory variables.

Figure 2.5. Estimated **forecasting regressions for cumulative bond market excess returns**. Forecasting horizons are on horizontal axes. Market returns are log returns (*not* annualized) on the value-weighted portfolio of TRACE bonds (upper panel) or the Barclays Aggregate corporate bond market index (lower panel). Explanatory variables are on vertical axes. See Appendix B.2 for the details on explanatory variables. Each point on a solid line on each chart is the OLS-estimate from a corresponding regression. Shaded areas around are two standard errors of estimates. The standard errors are heteroskedasticity and autocorrelation consistent estimates of Newey and West (1987). Each model also includes a constant (not reported). The sample is daily from Oct 4, 2004 to Dec 23, 2014.

My preferred measure of the EBP remains a significant predictor of bond market returns when the market is the Barclays Aggregate U.S. corporate bond index.[15] On Figure 2.5b, I present these estimated forecasting regressions. Both statistical and economic significance of the EBP still holds, moreover, here the EBP is a significant predictor on all horizons from several weeks to several months ahead. In contrast, the fitted spread is nowhere significant. The adjusted R^2 of return-forecasting regression at the 20-day horizon is 0.33. Appendix B.4 demonstrates that the predictive power of the EBP for market returns remains if I control for the VIX levels in returns-forecasting regressions. To sum up, out of all considered factors the EBP free from aggregate business risk and bond liquidity risk is the only significant in-sample predictor of cumulative corporate bond market returns 1–3 months ahead.

2.5.2 'Real-time' EBP as a Predictor of Market Returns

The EBP constructed and discussed in Sections 2.3 and 2.4 is the in-sample measure based on the entire dataset as of the end of 2016. A 'real-time' EBP might, in principle, be different from my full-sample measure because the whole historical path of the ADS index is re-estimated as new macroeconomic data become available (see Appendix B.2 for details). In this section, I estimate a 'real-time' EBP, show that it is not much different from the full-sample measure, and demonstrate that the two have similar predictive power for bond market returns.

Computation of a 'real-time' EBP is possible for all dates starting from the end of 2008; this is when the historical vintages of the ADS become available.[16] For every single day t in the sample, I cut my bond-day data at day t, take the ADS vintage as of t, and re-run Model

[15]The correlation of excess returns on Barclays index with excess returns on our TRACE portfolio is 0.79.
[16]See the Philadelphia Fed web-page: `https://goo.gl/mZJ5Sj`.

6 of Table 2.2 on this dataset to obtain the real-time measure of EBP denoted $EBP^{RT}(t)$.[17] I will denote observations in this time series $EBP^{RT}_{\tau}(t)$, where $\tau \leq t$.

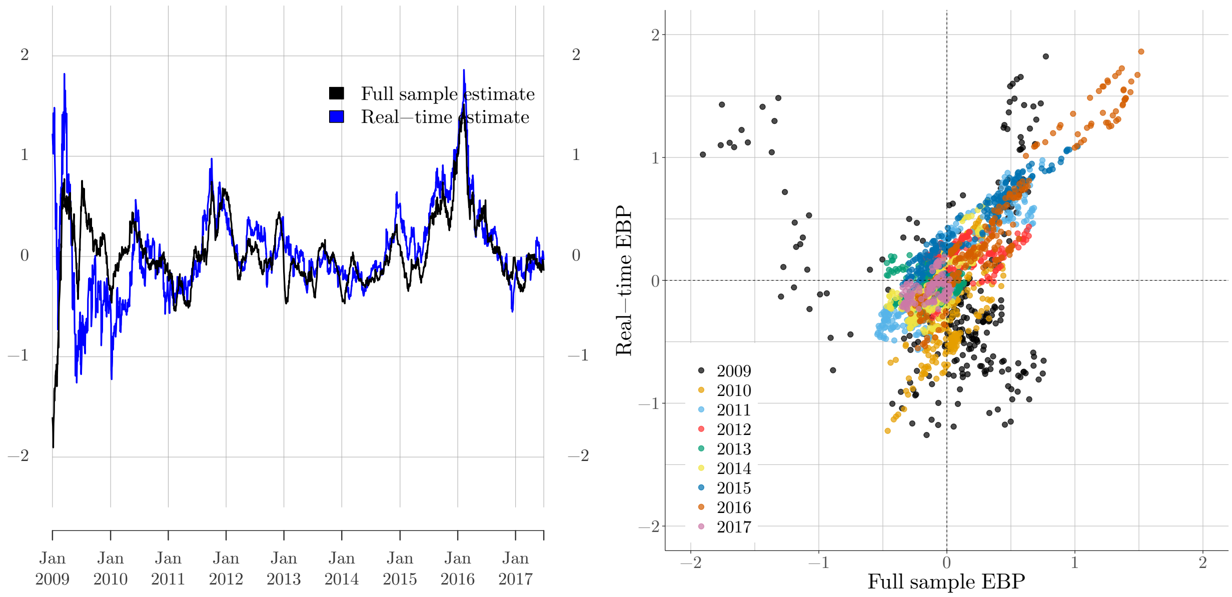


Figure 2.6. Real-time daily EBP measures computed with only aggregate activity data available on each estimation day, in comparison with full-sample EBP estimates (same as 'Model 6' on the bottom-left panel of Figure 2.2); time-series on the left and scatter plot on the right. Each daily observation of the real-time EBP is computed by re-estimating Model 6 of Table 2.2 for log spreads with a historical ADS vintage available on that particular day, and taking the latest EBP observation. Re-estimations are performed on expanding samples; each spans a period from Oct 4, 2004 to the estimation day.

Real-time EBPs turn out to be not much different from the full-sample EBP starting from the year 2010, as charts on Figure 2.6 demonstrate. These charts present a collection of the last points of real-time EBPs: $\left\{EBP^{RT}_{\tau=t}(t)\right\}$. Here, I estimate $EBP^{RT}(t)$ with samples always starting on Oct 4, 2004, and ending on the estimation day t. There are periods of time in 2009 when real-time EBPs differ considerably from the full-sample EBP_t estimate; otherwise, the real-time and the full-sample measures are close. Hence, we may expect that

[17] There is still one piece of information I use that could have not been available at day t, namely, accounting books used to compute the distance-to-default. I do not expect, however, real-time accounting books to diminish the explanatory power of the ADS index for credit spreads. Late dissemination of information about idiosyncratic credit risk would probably increase the loading on timely systematic business risk measure in explanatory regressions for credit spreads.

whatever valuable information EBP_t contains, we can extract it in real time, unless we are in some very volatile period as 2009 was.

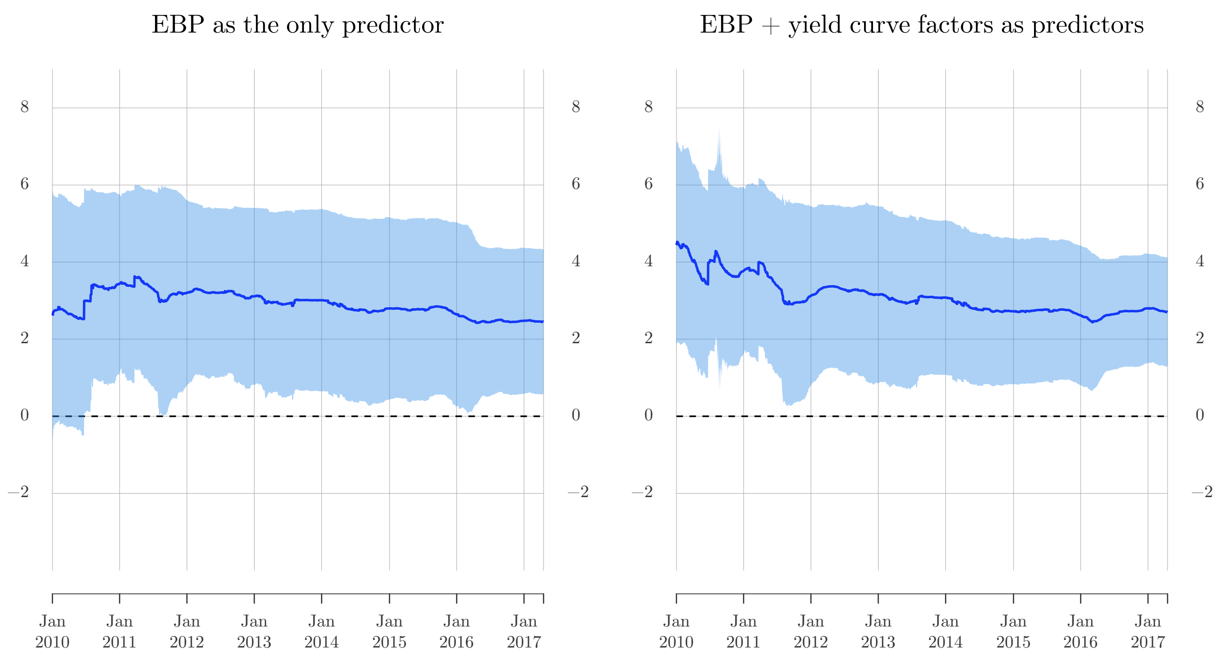


Figure 2.7. Coefficients on real-time EBP in cumulative excess corporate bond market return forecasting regressions. The dependent variable is the cumulative 50-day ahead log return on the value-weighted portfolio of TRACE bonds, *not* annualized. The left chart presents the estimates from the model with real-time EBP as the only predictor. The right chart presents the estimates from the model that also includes yield curve factors (level, slope, and curvature) and one lagged cumulative bond market return as predictors. The underlying samples expand from Oct 4, 2004 to estimation dates, which are on the horizontal axes of the charts. Lines are OLS-estimates of the coefficients on EBP from corresponding regressions. Shaded areas are two standard errors of estimates. The standard errors are heteroskedasticity and autocorrelation consistent estimates of Newey and West (1987).

It's important to check, however, whether the predictive power of the EBP for corporate bond market returns holds when the full-sample estimate is replaced in forecasting regressions with real-time estimates. I demonstrate in Figure 2.7 that, ever since 2010, real-time EBP has mostly been a significant predictor for excess bond market returns. Here, I re-estimate for each day t two forecasting models for 50-days ahead excess cumulative bond market returns with the real-time EBP on the right:

$$R_{\tau:\tau+50} = \alpha + \gamma EBP_{\tau}^{RT}(t) + \epsilon_{\tau+50},$$

$$R_{\tau:\tau+50} = \alpha + \beta R_{\tau-50:\tau} + \gamma_1 LVL_{\tau} + \gamma_2 SLP_{\tau} + \gamma_3 CRV_{\tau} + \gamma_4 EBP_{\tau}^{RT}(t) + \epsilon_{\tau+50}.$$

Figure 2.7 depicts estimated coefficients $\hat{\gamma}$ and $\hat{\gamma}_4$ and their confidence bounds for each estimation day t (on the horizontal axis). The left chart demonstrates that real-time EBP has significantly predicted excess bond market returns in-sample since 2010 in the univariate regression model. The right chart of Figure 2.7 indicates that this predictive power is not affected by the inclusion of additional yield curve factors in the model. Here, for almost all estimation days in 2010–2014, EBP is still a significant predictor of excess corporate bond returns 50 days ahead. A 10 b.p. rise in $EBP_{\tau=t}^{RT}(t)$ implies 25 to 40 b.p. extra excess cumulative bond market returns over $t : t + 50$ when t is in 2010–2014.

The analysis so far focused on in-sample predictability. Now I use the two models of this paragraph to investigate out-of-sample predictability of corporate bond market returns with the real-time EBP estimates. Figure 2.8 presents out-of-sample predictive accuracy tests of Diebold and Mariano (1995) in which my models are tested against the no-predictability benchmark (zero expected excess corporate bond market returns) on forecasting horizons from 1 to 90 days ahead. The forecasts are constructed for all trading days in 2010–2014. The null states that candidate models are as accurate as zero excess return forecasts in this period. As Figure 2.8 shows, the null is rejected in favour of out-of-sample return predictability on horizons shorter than 10 days and longer than 45 days ahead when the EBP is the only predictor in the model.

These real-time estimations confirm that the EBP contains useful information for forecasting excess corporate bond market returns. The cheaper corporate bonds are relative to risk-free counterparts today (controlling for firm-specific credit risk, bond-specific liquidity

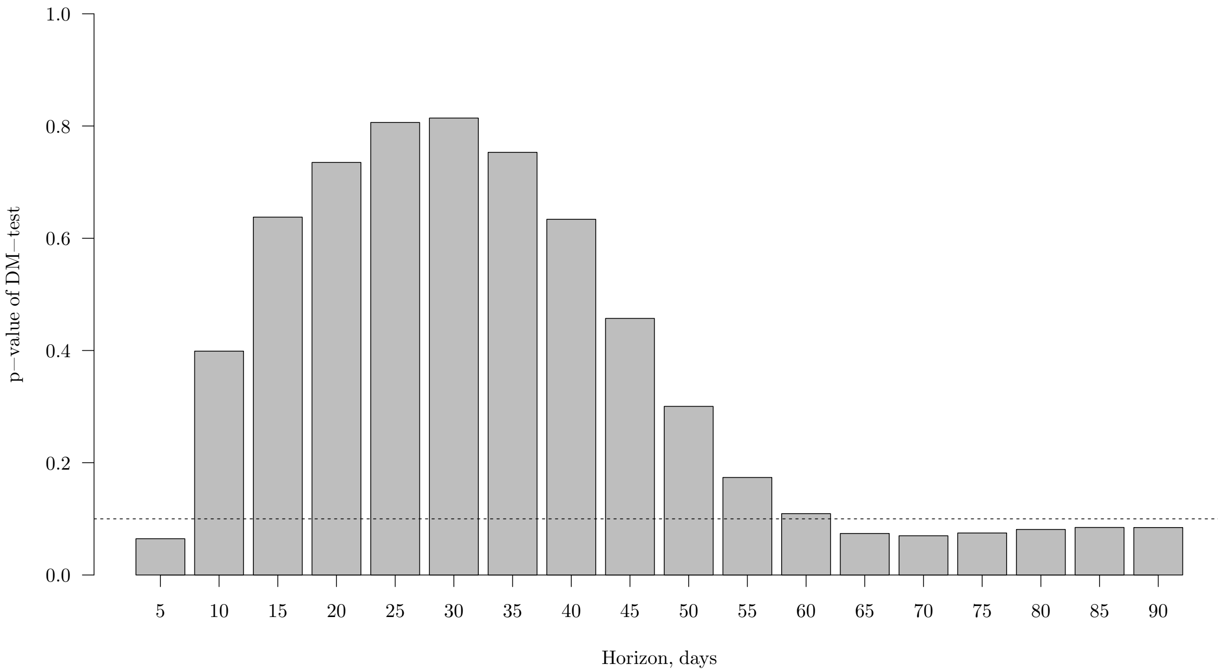


Figure 2.8. Diebold-Mariano (DM) out-of-sample predictive accuracy test of return-forecasting models relative to no-predictability (zero expected excess return) benchmark on different forecasting horizons. Candidate predictive models for cumulative corporate bond market all have the real-time EBP as the only predictor. The forecasting horizon is on the horizontal axis. The null is equal predictive accuracy with zero excess-return benchmark. The alternative hypothesis is greater out-of-sample predictive accuracy of a considered return forecasting model. P-values of the DM test are on the vertical axis. Values below 0.1 indicate rejection of the null at the 90% confidence level. The bond market is the Barclays IG portfolio of bonds. DM test statistics are computed using forecast errors for all trading days between Jan 4, 2010 and Dec 23, 2014.

risk, and aggregate business risk), the more they deliver on average over the next several months.

2.5.3 EBP and Other Corporate Bond Risk Factors

In this section, I demonstrate that the EBP is not explained by other corporate bond pricing factors, yet it improves their forecasting power with respect to diverse test portfolio returns. In particular, I compare the EBP to bond pricing factors derived by Bai et al. (2019) (referred to as 'BBW factors' herein). These factors are the 'market' factor, default risk factor (DRF), credit risk factor (CRF), and liquidity risk factor (LRF). These empirical factors are returns on factor-mimicking portfolios (see Appendix B.2 for details about the construction of the factors). Bai et al. (2019) demonstrated that their four factors explain the major portion of variation of bond returns for size and maturity decile portfolios.

I do not have access to the original time series of the BBW factors, so I re-estimate them on my sample using the methodology by Bai et al. (2019). They compute the factors on the sample of TRACE bonds over a comparable time frame (Jul 2002 – Dec 2014) at a monthly frequency with monthly portfolio rebalancing. I compute the BBW factors either as in the original work with monthly portfolio rebalancing ('monthly factors') or, as a robustness check, with daily rebalancing ('daily factors').

As Table 2.5 shows, the EBP is not strongly correlated with the bond risk factors, neither at the monthly nor the daily frequency. Likewise, the EBP is not linearly related to any stock market factor. The factors that are mildly correlated with the EBP on the monthly frequency are limited to bond credit risk and stock momentum factors. In the regression of daily EBP on BBW factors and a constant (not reported), none of the regressors has a significant coefficient, and the overall explanatory power of such a regression is low (adjusted R^2 is below 0.1). From this, I conclude that major empirical bond and stock pricing factors do not explain the time series variation of the EBP.

As Table 2.6 shows, the EBP does not add much to the BBW factors in explaining returns on Bai et al. (2019) test bond portfolios. Here, I consider monthly returns of size and maturity decile portfolios which I try to explain using candidate risk factors.[18] I also add industry portfolios to the analysis.[19] The 'Explanatory model' parts of Table 2.6 present R^2 from regressions of test portfolio returns on candidate risk factors. The columns titled BM and 4F refer to regressions with only the market factor and all the BBW factors correspondingly. The columns titled BM+ and 4F+ add the EBP as an explanatory variable to these baseline models. The market factor alone explains, on average, about 60-65% of variation of test

[18] As in Section 6 of the work by Bai et al. (2019) I also tested 25 maturity-size quintile portfolios, but all the subsequent results are qualitatively similar for them, so they are not reported. For the same reason I am not reporting results obtained on the daily frequency.

[19] Eleven industry portfolios based on two-digit NAICS codes of the issuing firms.

	EBP	BM	DRF	CRF	LRF	SM	SMB	HML
EBP								
BM	0.03							
DRF	-0.02	0.32***						
CRF	-0.09***	-0.30***	-0.03					
LRF	0.09***	0.24***	0.53***	-0.08***				
SM	-0.01	-0.28***	-0.01	0.37***	-0.02			
SMB	-0.01	-0.09***	-0.03	0.14***	-0.03	0.31***		
HML	0.00	-0.16***	-0.04*	0.14***	-0.05**	0.40***	0.09***	
UMD	-0.01	0.10***	-0.02	-0.15***	-0.01	-0.36***	-0.08***	-0.56***

(a) Correlation matrix of **risk factors on the daily frequency**. Underlying portfolio rebalancing in construction of the corporate bond risk factors is also daily. The sample starts on Mar 10, 2005, because first 100 days are needed to compute the first observation of the DRF factor.

	EBP	BM	DRF	CRF	LRF	SM	SMB	HML
EBP								
BM	0.15							
DRF	-0.06	0.50***						
CRF	-0.30***	-0.11	-0.28**					
LRF	-0.04	0.22**	0.59***	-0.18*				
SM	0.02	-0.10	-0.04	-0.12	0.06			
SMB	0.15	0.01	-0.11	-0.03	0.16	0.10		
HML	0.10	-0.21*	-0.20*	-0.16	-0.02	0.44***	-0.06	
UMD	-0.24**	-0.05	-0.03	0.10	0.00	-0.29***	0.08	-0.46***

(b) Correlation matrix of **risk factors on the monthly frequency**. Portfolio rebalancing frequency in construction of risk factors is monthly. EBP is monthly means of the daily series. The sample starts on Oct 2007, because first three years are needed to compute the first observation of the DRF factor.

Table 2.5. Correlations of the EBP with corporate bond risk factors from Bai et al. (2019), and stock market risk factors. The EBP is a full sample estimate from Model 6 of Table 2.2. The upper panel uses daily time series of the EBP; lower panel uses monthly averages of the EBP. BM stands for excess returns on the aggregate bond market index (Barclays IG), DRF for the default risk factor, CRF for the credit risk factor, and LRF for the liquidity risk factor. Construction methodology for the risk factors is similar to Bai et al. (2019), the details are provided in Appendix B.2. The difference between the upper and the lower panels is a frequency of portfolio rebalancing for the construction of bond risk factors. The last four rows of correlation matrices refer to the Fama-French stock market risk factors: SM is excess market return, SMB is the small-minus-big factor, HML is the high-minus-low factor, and UMD is the momentum factor. Significance code: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

portfolio returns. Three additional risk factors, DRF, CRF, and LRF, add 12–15% to the R^2 of explanatory models on average. The EBP adds to that average virtually nothing. Based on these findings, I conclude that the EBP does not explain residual bond returns that are not explained by the BBW factors.

	Explanatory model				Forecasting model			
	BM	4F	BM+	4F+	BM	4F	BM+	4F+
D1	0.31	0.61	0.34	0.62	0.34	0.33	0.36	0.33
D2	0.52	0.81	0.52	0.82	0.24	0.26	0.34	0.36
D3	0.54	0.75	0.54	0.75	0.22	0.21	0.30	0.27
D4	0.67	0.83	0.67	0.83	0.15	0.17	0.26	0.25
D5	0.68	0.84	0.68	0.84	0.14	0.14	0.24	0.24
D6	0.73	0.85	0.73	0.85	0.11	0.12	0.23	0.23
D7	0.76	0.87	0.75	0.87	0.07	0.08	0.19	0.18
D8	0.81	0.90	0.81	0.90	0.05	0.09	0.20	0.20
D9	0.81	0.90	0.81	0.90	0.03	0.07	0.17	0.17
D10	0.73	0.86	0.74	0.86	0.01	0.02	0.12	0.11
Average	0.66	0.82	0.66	0.82	0.14	0.15	0.24	0.23

(a) Maturity decile portfolios (D1 – shortest, D10 – longest).

	Explanatory model				Forecasting model			
	BM	4F	BM+	4F+	BM	4F	BM+	4F+
D1	0.41	0.79	0.42	0.79	0.22	0.17	0.24	0.18
D2	0.53	0.85	0.53	0.85	0.21	0.21	0.26	0.23
D3	0.66	0.85	0.66	0.85	0.15	0.17	0.24	0.23
D4	0.68	0.86	0.68	0.86	0.11	0.10	0.19	0.17
D5	0.69	0.83	0.68	0.83	0.10	0.09	0.18	0.15
D6	0.77	0.85	0.77	0.85	0.08	0.09	0.20	0.19
D7	0.80	0.90	0.80	0.90	0.06	0.06	0.19	0.18
D8	0.79	0.87	0.79	0.87	0.05	0.04	0.15	0.13
D9	0.86	0.91	0.86	0.91	0.03	0.07	0.19	0.20
D10	0.86	0.89	0.86	0.89	0.01	0.08	0.16	0.19
Average	0.71	0.86	0.71	0.86	0.10	0.11	0.20	0.19

(b) Size decile portfolios (D1 – smallest, D10 – largest).

	Explanatory model				Forecasting model			
	BM	4F	BM+	4F+	BM	4F	BM+	4F+
Mining, Quarrying, and Oil and Gas Extraction	0.64	0.73	0.64	0.73	0.11	0.04	0.17	0.10
Utilities	0.69	0.84	0.68	0.84	0.15	0.16	0.24	0.23
Construction	0.31	0.42	0.31	0.42	0.05	0.15	0.12	0.17
Manufacturing	0.76	0.91	0.76	0.91	0.07	0.09	0.20	0.18
Wholesale Trade	0.77	0.85	0.77	0.86	0.09	0.10	0.19	0.16
Retail Trade	0.75	0.84	0.75	0.84	0.05	0.07	0.16	0.16
Transportation and Warehousing	0.65	0.82	0.65	0.82	0.12	0.08	0.20	0.17
Information	0.78	0.88	0.78	0.88	0.07	0.06	0.16	0.14
Professional, Scientific, and Technical Services	0.23	0.39	0.23	0.40	0.03	0.09	0.02	0.08
Administrative and Support etc. Services	0.42	0.54	0.42	0.54	0.05	0.07	0.12	0.10
Health Care and Social Assistance	0.33	0.59	0.33	0.59	0.06	0.08	0.17	0.19
Accommodation and Food Services	0.66	0.76	0.66	0.76	0.08	0.08	0.18	0.17
Average	0.58	0.72	0.58	0.72	0.08	0.09	0.16	0.15

(c) Industry portfolios (2-digit NAICS codes).

Table 2.6. Adjusted R^2 of explanatory and forecasting regressions for monthly returns on size decile, maturity decile, and industry portfolios. In explanatory models returns and risk factors are contemporaneous, in forecasting models returns are one month ahead. The sample is monthly from Oct 2007 to Dec 2014. Four alternative models are considered: BM has the bond market risk factor as the only explanatory variable, 4F has DRF, CRF, and LRF factors in addition, BM+ has the market factor and the EBP, and 4F+ has four aforementioned factors and the EBP.

However, when I run forecasting regressions for one-month ahead test portfolio returns, the EBP does much better than the BBW factors. Note first in the 'Forecasting model' parts of Table 2.6 that the market factor alone forecasts returns almost as good as the full Bai et al. (2019) four-factor model. This result holds for all industry portfolios and most size and maturity portfolios. That is why I use the 'BM' columns as benchmarks for return-forecasting regressions with the EBP added. The EBP improves the forecasting power of return-forecasting regressions across the board. Returns on all size, maturity, and industry portfolios are better forecasted once the EBP is included in the forecasting regressions. In addition, the increase in R^2 between BM and BM+ columns is pretty uniform across test portfolios. Hence, the strong forecasting power of the EBP is hardly due to any specific size, maturity, or industry group of bonds.

	Full sample			Excl. Sep-Dec'08		
	Maturity	Size	Industry	Maturity	Size	Industry
BMarket	8	8	12	7	6	7
EBP	10	10	11	9	10	12
Memo: # of portf.	10	10	12	10	10	12

Table 2.7. Significance of explanatory variables in BM+ return-forecasting models from Table 2.6 at the 95%-level . A number in row i of column j shows for how many portfolios of type j factor i is a significant one-month ahead predictor of returns. Total number of portfolios of type j is given in the last line of the table. The left three columns build forecasting models over the full sample (Oct 2007 – Dec 2014), while the right three columns drop the period from Sep 2008 to Dec 2008 inclusive.

Next, I check that the forecasting power of the EBP does not hinge on extreme return observations of the end of 2008. In Table 2.7, I compare for how many test portfolios of each type the coefficient on the EBP is significant in return forecasting-regressions both for the entire sample and for the sample with observations from Sep 2008 to Dec 2008 removed.[20] Table 2.7 demonstrates that the EBP remains a significant predictor of test portfolio returns at the 95% level, even with four extreme monthly observations removed. Removing years 2008 and 2009 completely makes this monthly time series very short, but even in this case

[20]These are the months with very low market returns and very high EBPs.

(not reported), the EBP remains a significant predictor of returns at a 90% confidence level for most portfolios and for more portfolios than the market factor.

The logic behind the results of this section follows. Given the predictive power of the EBP for bond market returns discussed in Section 2.5, one should expect the EBP to forecast also whatever is strongly correlated with the market. As Table 2.6 shows, the market factor explains the major portion of variation of a broad range of test portfolio returns. Hence, one should expect the forecasting regressions in Table 2.7 to perform well, as they indeed do. It is important, though, that this result is not attainable with other bond pricing factors. The EBP outperforms DRF, LRF, and CRF factors in forecasting bond returns. In the next paragraph, I demonstrate how the predictive power of the EBP can be used to construct an investment strategy that outperforms the corporate bond market.

2.5.4 Corporate Bond Market-timing Strategy

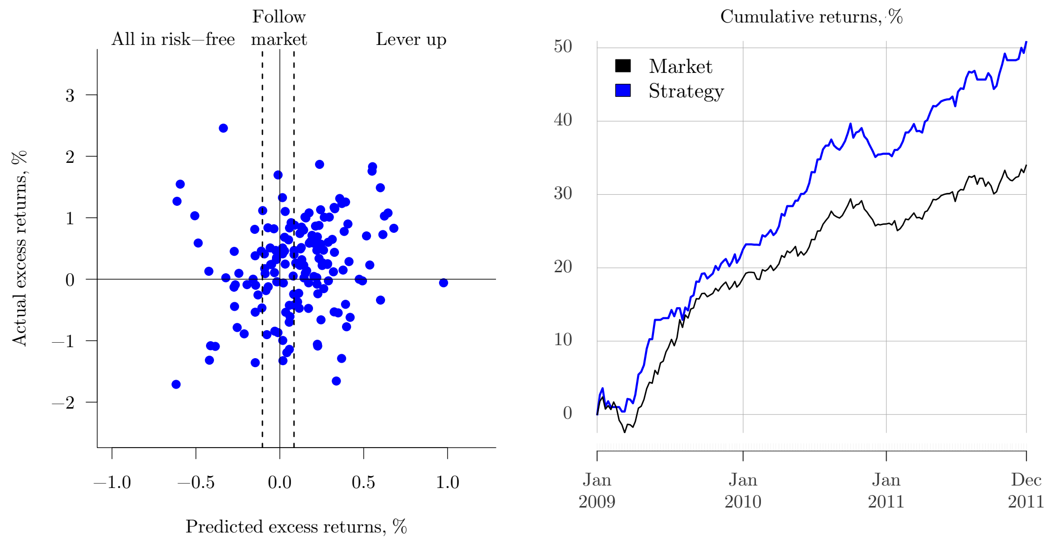
I use the predictive power of the EBP for corporate bond market returns to design a market-timing strategy that delivers risk-return characteristics superior to the buy-and-hold the market strategy. My strategy uses only one risky instrument: the Barclays Aggregate U.S. corporate bond market index (investable; several replicating ETFs are available). The strategy consists of making one-week ahead forecasts of corporate bond market excess returns using recent observations of the EBP, fitted GZ spread, yield curve factors, and market returns. Based on these forecasts, an investor who has an amount of money W under management at the end of week t can take one of the following three positions for the week $t+1$:

- stay away from the corporate bond market and invest W in risk-free securities only (when low returns are forecasted);
- follow the market and invest W in the index ETF (when the model provides no clear signal about future returns);

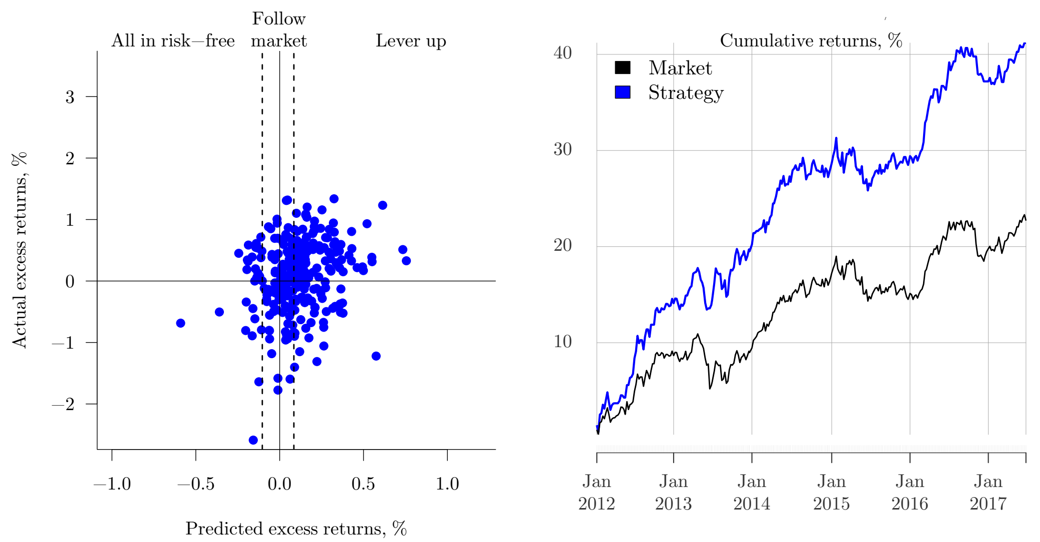
– borrow a certain fraction α of W at the risk-free rate, and invest $(1+\alpha)W$ in the index ETF (when high positive returns are forecasted).

The forecasting model builds upon the results of Sections 2.5.1–2.5.3. The left-hand side variable is the *weekly* corporate bond market excess returns. The right-hand side variables are the five latest *daily* observations of the EBP, fitted GZ spread, three yield curve factors, and daily corporate bond market returns *one week prior to return observations*. Hence, there are 30 explanatory variables in total; selection among them is done by running LASSO estimations. The model is re-estimated every week w (using 'real-time' estimates of the EBP and fitted GZ spread of Section 2.5.2), and the LASSO penalty parameter λ is selected to minimize the root mean squared error (RMSE) of the out-of-sample forecasts with the 'leave-one-out' cross validation. Once the model is estimated, the forecast for the next week $w+1$ is made using daily observations of predictors on week w.

The boundaries of the 'inaction region' in terms of predicted returns (when the investor simply holds W in the corporate bond market ETF) and the leverage ratio α are selected over the training sample, which is years 2009–2011. The selection problem is solved by maximizing the Sharpe ratio of the market-timing strategy on the training sample. The optimization is constrained, the lower bound of the inaction region is required to be negative, the upper bound positive, and $0 \leqslant \alpha \leqslant 0.5$. The left chart of Figure 2.9a presents the out-of-sample one-week ahead forecasts of market excess returns vis-a-vis actual excess returns. The two are significantly correlated: the correlation coefficient is 0.26, the regression coefficient is 0.96 (in the regression of actual returns on predicted ones), and both are significant at the 1%-level. Maximizing the Sharpe ratio yields $\alpha = 0.5$, the lower bound of the inaction region of -0.12%, and the upper bound of 0.06% (of predicted weekly market excess return). Table 2.8 and the right chart of Figure 2.9a show how the market-timing strategy performs on the training sample. It delivers 50% cumulative return over the three years (1.5 times more than the market) with a weekly Sharpe ratio of 0.37 (1.3 times higher than the market).



(a) Training sample: 2009 – 2011. Expected return bounds (vertical dashed lines) that determine investments for a week ahead (actions that are taken are given at the top of the left chart) are selected to maximize the Sharpe ratio of the strategy. The leverage ratio is 0.5.



(b) Testing sample: 2012 – 2017. Expected return bounds (vertical dashed lines) and 0.5 leverage ratio are as determined on the training sample.

Figure 2.9. Out-of-sample forecasts of corporate bond market excess returns vis-a-vis actual returns, **and comparative performance of the market timing strategy** based on these forecasts. The return forecasting model is a 'leave-one-out' cross-validated LASSO regression with the penalty parameter selected to minimize the out-of-sample RMSE at each re-estimation date, which is weekly. The dependent variable is weekly excess corporate bond market returns. The regressors are 5 latest *daily* observations of excess returns, yield curve factors (level, slope, and curvature), fitted GZ spreads, and the EBP, *all one week prior to returns observations.* The EBP and the GZ spread are real-time expanding sample estimates of Section 2.5.2. Transaction costs are not accounted for.

	Train (2009–2011)		Test (2012–2017)	
	Market	Strategy	Market	Strategy
Mean excess return	0.22	0.32	0.08	0.14
Standard devaition	0.76	0.88	0.59	0.71
Sharpe ratio	0.29	0.37	0.13	0.20
Information ratio		0.23		0.21
Max. excess return	2.46	2.80	1.34	2.01
Min. excess return	-1.71	-2.48	-2.59	-2.10

Table 2.8. Comparative performance of the corporate bond market (Barclays Aggregate U.S. corporate bond index, investable) **and a proposed market timing strategy**. Returns and standard deviations are in *% per week*, not annualized. The strategy consists in making 1-week ahead forecasts of market excess returns and taking positions in a market ETF based on these forecasts. Three options are available: invest all in risk-free bonds (low expected excess returns), follow the market (mediocre expected returns), lever up and invest more in the market (high expected returns). Separation bounds in terms of expected returns are determined by maximizing the strategy Sharpe ratio of the strategy over the training sample (see Figure 2.9 also). The leverage ratio is 0.5, meaning that 50% of the accumulated asset value is borrowed for one week and invested in the market whenever the strategy prescribes to lever up. The return-forecasting model is a 'leave-one-out' cross-validated LASSO regression with the penalty parameter selected to minimize the out-of-sample RMSE at each re-estimation date, which is weekly. The dependent variable is weekly excess corporate bond market returns. The regressors are 5 latest *daily* observations of excess returns, yield curve factors (level, slope, and curvature), fitted GZ spreads, the EBP, and a month dummy *all one week prior to return observations.* The EBP and the GZ spread are real time expanding sample estimates of Section 2.5.2. The information ratio in the table is relative to the corporate bond market returns. Transaction costs are not accounted for.

As Figure 2.9b and Table 2.8 demonstrate, the strategy performs equally well on the testing sample, which is years 2012–2014 (with $\alpha = 0.5$ and inaction region bounds fixed at the values found on the training sample). Out-of-sample forecasts of market excess returns are again strongly correlated with actual returns; the correlation coefficient is 0.22, and the regression coefficient is 0.94, while both are significant at the 1%-level. Out of 155 weeks in the testing sample, an investor follows the market for 49 weeks, levers up for 93 weeks, and stays away from the market for 13 weeks. The strategy increases both mean weekly returns and the Sharpe ratio by roughly one-half relative to the buy-and-hold market strategy.[21]

[21] Transaction costs are not accounted for, but given that the strategy uses only one instrument, which is traded on the market, they will not considerably affect the results.

Cumulative returns of the strategy over the three testing years is 28% compared to 16% of the corporate bond market index.

2.6 Conclusion

In this chapter, I explore the forecasting power of the aggregate corporate bond risk premium (EBP) with respect to the business cycle and corporate bond market returns. Unlike the closest study Gilchrist and Zakrajšek (2012), that defines the EBP as the portion of credit spread not explained by firm-specific credit risk, I additionally project spreads on bond-specific liquidity risk and economy-wide business risk. I do so using daily data constructed from tick-by-tick high-frequency data, while the literature works so far with historical monthly data.

The chapter demonstrates that the forecasting power of the EBP for future economic activity depends on whether the EBP contains information about contemporaneous liquidity and aggregate business risks. The residual spread that is free from only corporate credit risk indeed forecasts activity, but this forecasting power mostly hinges on bond liquidity and aggregate business cycle states. The latter two are readily measurable with daily frequency. Once this information is taken away from credit spreads, both the statistical and the economic significance of the residual for the forecasts of macroeconomic activity reduces a lot.

This residual spread, however, forecasts corporate bond market returns, unlike the EBP correlated with bond liquidity and aggregate business risks. The forecasting power is robust to different definitions of the bond market portfolio and to different estimation windows. I demonstrate that major stock and bond risk factors, including contemporaneous bond market returns per se, do not explain the time series variation of my risk premium measure. Moreover, its forecasting power is not concentrated in any particular size, maturity, or industry portfolio; the risk premium improves forecasts of corporate bond portfolios across the board.

One can profit from the forecasting power of the residual spread by investing according to the strategy designed to time the corporate bond market. The chapter constructs the forecasting model for the corporate bond market excess returns that successfully forecasts returns out-of-sample. The strategy consists of staying away from the market when low negative returns are forecasted and levering up when high positive returns are forecasted; otherwise, an investor just follows the market. The strategy is implemented with only one risky instrument, an aggregate corporate bond market ETF, and delivers risk-adjusted returns 50% higher than the buy-and-hold market strategy.

Chapter 3

(In)frequently Traded Corporate Bonds[1]

3.1 Introduction

Corporate bonds tend to trade actively on the secondary market for the first few months up to two years after issuance while they settle into the most-desired portfolios, and afterward, the trading thins out as many bonds are held to maturity, redemption, or a credit-default event. The early empirical literature on corporate bonds, e.g., Alexander, Edwards, and Ferri (2000) documented this anecdotal evidence. Nowadays, as the comprehensive TRACE data on corporate bonds trading has been available for more than a decade, it turns out not all bonds follow the conventional wisdom and there are notable and numerous exceptions from that rule above. Figure 3.1 shows a corporate bond that experiences substantial and long-lasting swings in trading activity sufficiently long after its issuance. We document that roughly 25% of all plain-vanilla fixed-coupon bonds stand out from the conventional wisdom and experience swings in trading activity in our sample period from January 2005 to July 2017. We call these bonds (in)frequently traded or the (I)TBs.

[1]This chapter is based on the paper co-authored with Artem Neklyudov, former Assistant Professor of Finance at the University of Lausanne, currently at Lancaster University Management School.

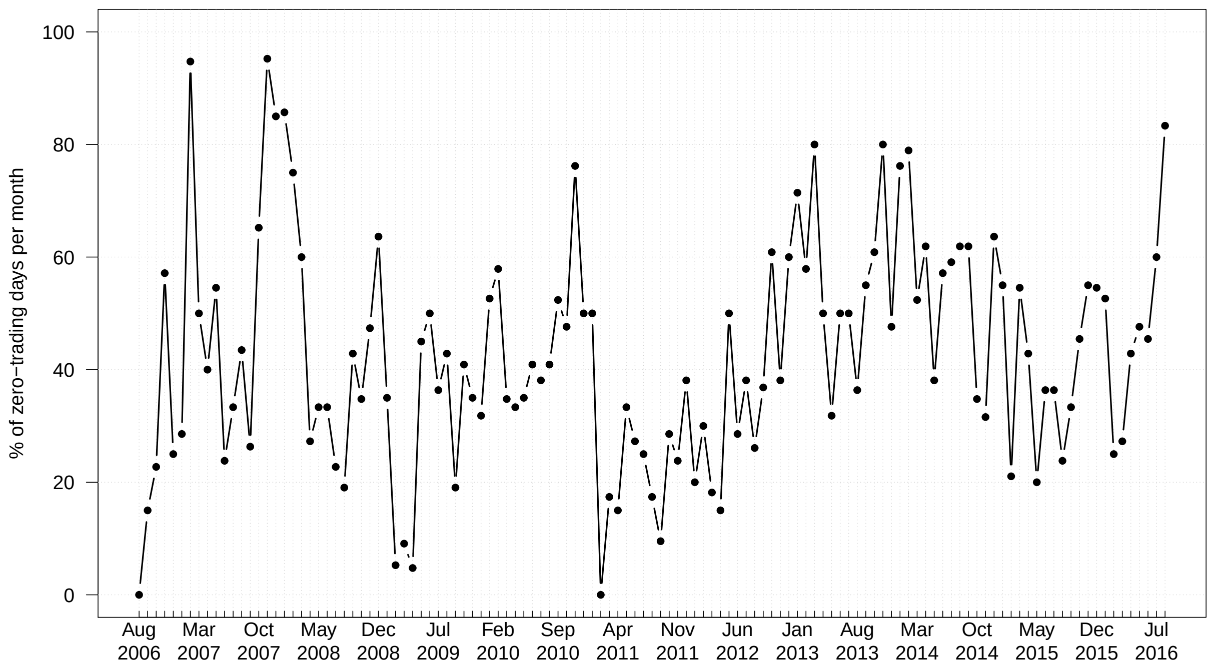


Figure 3.1. Fraction of zero-trading days per month for the Credit Suisse senior unsecured USD-denominated 500 mln USD 10Y 5.85% bond issued in Aug'06; CUSIP: 225434CJ6.

Surprisingly, the (in)frequently traded bonds are almost indistinguishable from all other plain-vanilla fixed-coupon bonds in major bond characteristics, including the issue size, average age, credit quality, etc. One cannot recover the information contained in the trading activity waves from headline bond characteristics.

Moreover, we find substantial excess returns associated with changes in bond trading frequency, but only post-crisis and only for the subsample of (in)frequently traded bonds. Figure 3.2 compares excess returns for (in)frequently traded bonds and all other bonds in our sample. The returns of the (I)TBs that move to states with higher trading frequency are about 12 basis points per month higher compared to the (I)TBs that stay in the same trading frequency state. We show that the exposure to Bai et al. (2019) risk factors does not explain these returns. Abnormal excess returns of the (I)TBs jumping to higher trading frequency states are of the same magnitude and statistically significant.

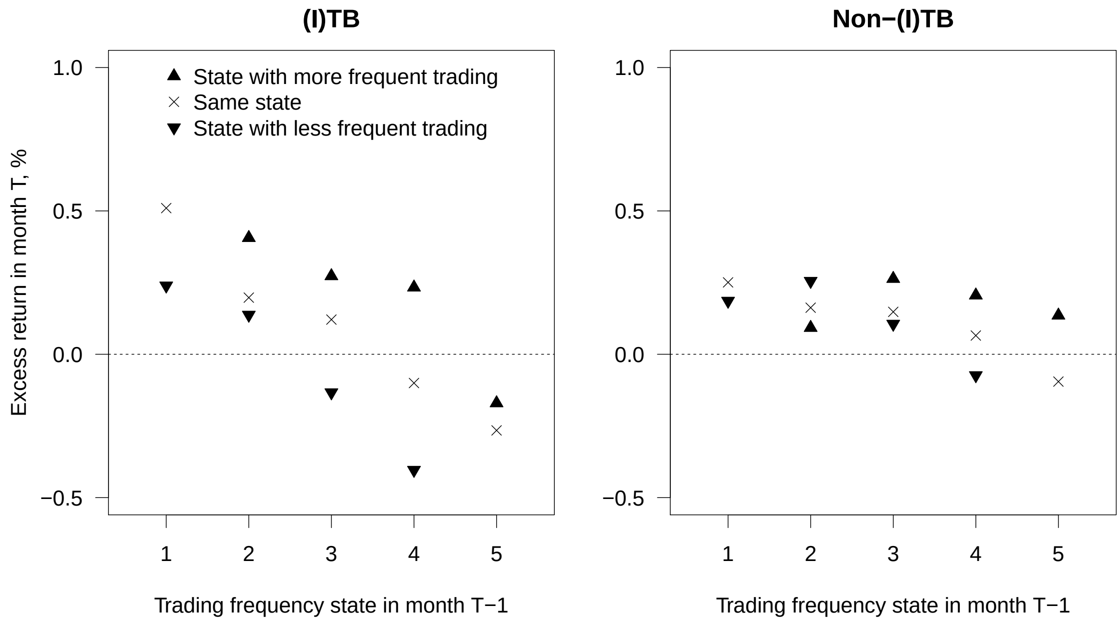


Figure 3.2. Mean excess returns and jumps between trading frequency states. State 1 is the state with the most frequent trading, and state 5 is the state with the least frequent trading. The cross represents the case when the bond stays in the same trading frequency state, the triangle pointed up represents a jump to a more frequent trading state (from 3 to 2, for example), and the triangle pointed down represents a jump towards less frequent trading (from 3 to 4, for example). Excess returns are returns above the 3-month T-Bill rate.

We document substantial differences between the (I)TBs and all other bonds in the structure of their trade flow and institutional ownership which as well sheds light on the nature of trading frequency changes of the (I)TBs. We found that the (I)TBs are more likely to be owned by mutual funds. Remarkably, there's a relatively constant number of funds that hold an (I)TB in any trading frequency state, unlike a Non-(I)TB that is held by substantially fewer funds when it trades infrequently. We also demonstrate that in the (I)TBs, compared to the rest of the sample, higher volumes are traded via small trades (less than 100'000 USD), and it takes more days to trade the same additional volume in small trades in the (I)TBs. It turns out that (I)TBs trading frequency goes up simultaneously with increases in net *purchases* by mutual funds and higher *sell* volumes in small trades by some investors. So, we link the waves of trading activity in the (I)TBs to mutual fund rebalancing. We also show that time-varying issue and issuer characteristics explain only a tiny portion of the variation of changes in corporate bond trading frequencies. From this, we conclude that mutual funds rebalancing that drives changes in trading activity is arguably not due to public corporate news.

Given that trading activity of (in)frequently traded bonds is related to demand from mutual funds, especially in the post-crisis period, one might expect that the link between trading frequency jumps and excess returns is due to the market impact of mutual fund purchases. We find some support for this view in the data, but institutional flows per se do not fully explain excess returns of the (I)TBs; the latter remains a puzzle.

We tend to think that the impact of trading frequency jumps on returns emerges from the interplay of higher demand from mutual funds, lower bond inventories among broker-dealers, and the desire of some smaller investors (supposedly, hedge funds) to take profit from cash corporate bond positions they have established in the wake of crisis sell-off. As documented in Dick-Nielsen and Rossi (2018), dealers prefer to keep low bond inventories post-crisis. So, the demand from mutual funds for the (I)TBs is not satisfied immediately as it takes time for dealers to accumulate positions and for investors to trade to their desired allocations. Smaller investors, who sell the (I)TBs, are likely to sell in small volumes, precisely as we observe in the data. Smaller trades tend to have the highest price impact in corporate bonds, as shown by Edwards, Harris, and Piwowar (2007b), and they contribute to excess returns of the (I)TBs. Since we do not observe dealer inventories and hedge funds positions in corporate bonds, we cannot test the described mechanism directly. However, much indirect evidence we present in this chapter is consistent with such an explanation.

To our knowledge, ours is the first work to look closely at the bond-by-bond variation in trading activity. Most empirical studies of corporate bond markets document that bonds trade only several times per day, and most bonds trade less than once a month (e.g., Edwards et al. (2007b), Bessembinder, Maxwell, and Venkataraman (2006)). We focus on sudden changes in trading activity. Our trading frequency measure is weakly correlated with changes in trading volume. Trading in corporate bonds is often pre-arranged. Harris (2015) documents that more than 40% of all trades in corporate bonds are riskless-principal trades. Large volumes may be traded within one business day and will not affect the waves of trading activity we analyze. Trading frequency is only weakly related to illiquidity measures either (e.g., the Bao et al. (2011) measure), and the relationship is weaker for the (I)TBs than for other bonds in our sample. Hence, this chapter extends beyond the existing discussion of corporate bond illiquidity and its impact on bond prices.

The chapter is organized as follows. Section 3.2 describes the data and the measure of trading frequency we use. In Section 3.3 we define (in)frequently traded bonds, document the differences between the (I)TBs and the rest of the sample in trade flows and mutual fund holdings, and attempt to explain monthly changes in trading frequency with institutional flows into the (I)TBs. In Section 3.4 we demonstrate that public news about issuers and issues do not drive changes in bond trading frequencies. Section 3.5 explores the relationship between bond trading frequency, returns, mutual fund holdings, trade flows, and exposure to corporate bond risk factors. Section 3.6 concludes.

3.2 Data and measurements

Corporate bonds in the U.S. are traded primarily on the OTC market, and trades are reported to the FINRA's Trade Reporting and Compliance Engine (TRACE). We use Enhanced TRACE data (contain uncapped volume records) available through WRDS in our study. Our sample consists of 'plain vanilla' corporate bonds only: unsecured fixed-coupon or zero-coupon bonds nominated in USD with the most typical coupon schedules and quoting conventions. We aggregate tick-by-tick TRACE data to the monthly frequency keeping in the sample all months when an outstanding bond *was not traded*. Volume is assumed 0 and prices missing (`NA`) for such bond-months. The sample consists of about 940 thousand bond-month observations covering approximately 14 thousand bonds issued by 2.6 thousand firms and traded for at least two days between Jan 1, 2005, and Jun 30, 2017. Roughly 25%

of bond-month observations refer to months when the bonds were not traded at all. We present the details on sample selection and data cleaning in Appendix C.1.

We obtain individual bond characteristics from the Mergent Fixed Income Securities Database (FISD) also available through WRDS. Besides, we use two pieces of data on institutional trading of corporate bonds. The transactions of insurance companies are reported to NAIC and are also available via Mergent FISD. For mutual fund transactions, we scrape the data from the SEC N-Q forms submitted by SEC-registered funds and available through SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. N-Q forms contain all mutual fund holdings; we focus on corporate bond holdings only. Changes in holdings represent net purchases by mutual funds in the reporting period. We describe the recovery of holdings from scraped textual data in Appendix C.2. As Table 3.1 shows, we recover mutual fund holdings for about 12 thousand out of 14 thousand bonds of the original sample; they cover about 740 thousand out of 940 thousand initial bond-month observations.

	Full sample	Subsample (SEC NQ)
Bond issues		
Unique securities	14,234	11,796
of them, identified as (I)TB	3,884	3,721
Bond-month observations		
Bond-month obs. (incl. non-traded)	938,229	736,514
of them, with identified returns	362,358	347,812
of them, identified as (I)TB	170,803	164,590

Table 3.1. Full sample and subsamples with identified mutual fund holdings and returns. For details on sample construction see Appendix C.1 and C.2.

In Chapter 3.5 we work with bond returns that are recognized using Bai et al. (2019) approach. First, we calculate volume-weighted daily (dirty) prices from the tick-by-tick TRACE data. Then, we calculate monthly returns if there are days with trades within five last business days of two consecutive months, or (if the first condition is false) in the first five and in the last five business days of a given month. In the first case, we use the latest volume-weighted daily prices of consecutive months to compute returns; in the second case, we use the earliest and the latest volume-weighted daily prices of a given month. Monthly returns in this study are total returns and contain coupon payments if there are any. Our return recognition approach results in about 360 thousand bond-month observations with recognized returns, which is roughly 40% of the original bond-month sample. Remarkably, about 96% of observations with recognized returns have identified mutual fund holdings.

In this chapter, we focus on the frequency of corporate bond trading. To measure the trading infrequency of bond i in month t we construct the fraction of zero-trading business days within that month, Z_{it}. Assume there are D_t trading days in a month t, and the bond i was only traded $\bar{D}_t \leqslant D_t$ of them.[2] Then

$$Z_{it} = 100 \cdot \left(1 - \frac{\bar{D}_t}{D_t}\right).$$

Hence, if the bond is traded every business day in month t, then $Z_t = 0$; if the bond is not traded at all in month t, then $Z_t = 100$. Z_t is the measure of trading infrequency we use throughout this chapter.

In a detailed study of trading cost and price impact proxies for corporate bonds Schestag et al. (2016) document that Z measure does not relate strongly to trading cost and price impact proxies both in the cross-section and in the time series. We find a similar pattern in our sample. In the pooled data the correlation coefficient between Z_{it} and, for instance, Bao et al. (2011) illiquidity measure is significant but small: pre-2008 crisis it stands at 0.08, post-crisis – at 0.13; the R^2 is less than 1.5%. In first differences, the correlation coefficients are twice smaller. Trading infrequency measure Z provides a different perspective on bond trading properties than typical illiquidity measures. In the next chapter, we also demonstrate that the relationship between illiquidity and trading infrequency differs across subsamples of the data.

Z is not correlated with total trading volume in levels: in the pooled data the correlation coefficient is statistically indifferent from zero. It comes as no surprise given the extent of pre-arranged trading in the corporate bond market. According to market participants, it

[2] We count a day in $\bar{D}_t$ if there is at least one trade of the bond on that day, regardless of the total trading volume.

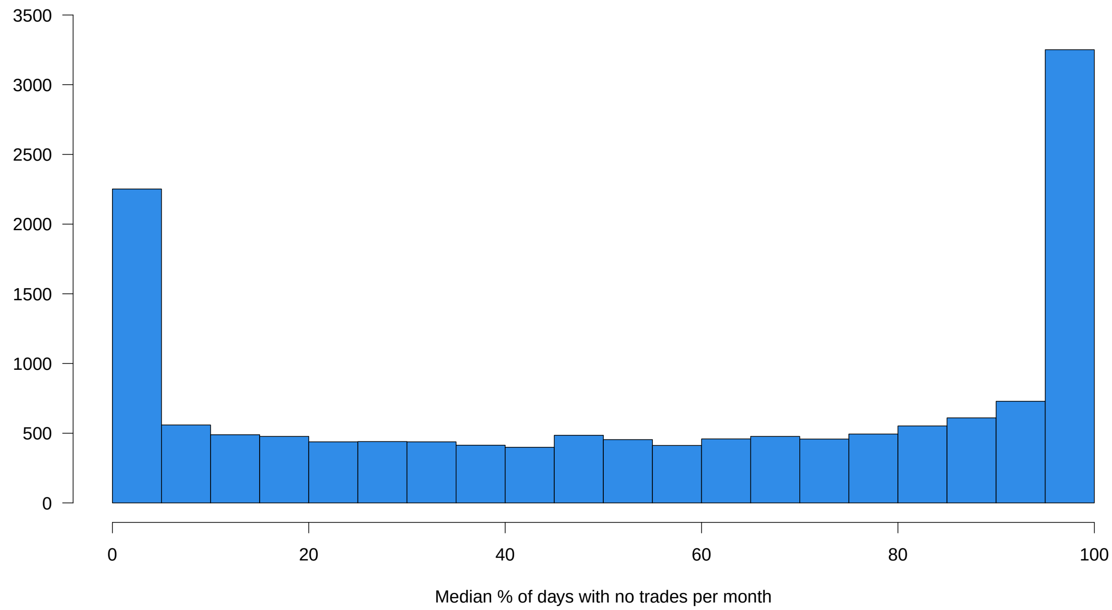


Figure 3.3. Distribution of bonds by a median fraction of zero trading days per month. The vertical axis counts bond issues in each bin. There are 14287 bonds in the sample. For each bond, the median is taken over its lifespan (defined as the time between dated date and maturity) that falls between Jan 1, 2005, and Jun 30, 2017.

takes time to discuss and prepare big trades, but once everything is set the execution occurs within one day. In the data, we indeed observe a high number of bond-months with very high volumes but very few trading days (hence, high Z). The relationship between *changes* in Z and the trading volume is more pronounced: volume increases are associated with decreases in Z. Interestingly, this relationship has different numerical properties in different subsamples of the data, we discuss it in more details in the next chapter.

Figure 3.3 plots the histogram of Z_i^{median} across bonds. It has two pronounced modes in the tails: there are about 2300 issues in the sample with Z^{median} below 5% (less than 5% non-trading days in the median month) and a thousand more issues with Z^{median} above 95%. The remaining mass of issues is almost uniformly distributed between the two tails.

To study how the distribution of Z across bonds changes over time, we partition the domain of Z (from 0 to 100%) into five intervals of equal length: 0 to 20% being the first one, 80 to 100% – the fifth one. We refer to these intervals as 'trading frequency groups'. The first group, we call it G1 (Z between 0 and 20), consists of the bonds that traded *at*

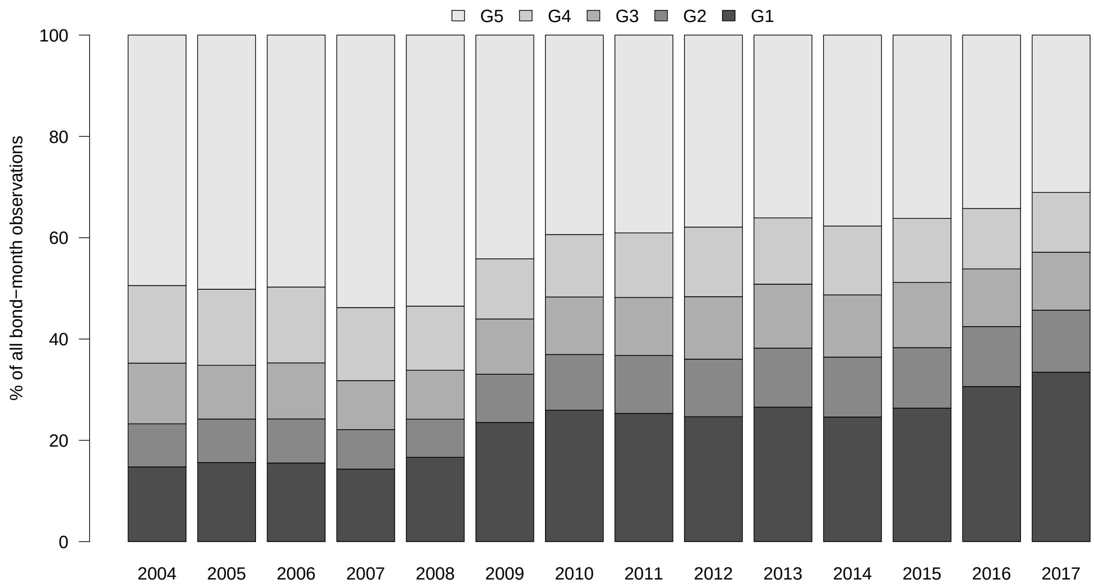


Figure 3.4. Distribution of bonds by trading frequency groups, per year. The histogram presents the number of bond-month observations in a given trading frequency group in a given year as % of the total number of bond-month observations in that year.

least four out of five trading days in each week on average in a month. These are frequently traded bonds. The last group, G5 (Z between 80 and 100), consists of the bonds that traded *at most one out of five trading days* in each week on average in a month. These are rarely traded bonds.[3] Figure 3.4 plots the distribution of bonds across trading frequency groups over time. Intermediate groups, G2–G4, contain about 35% of bond-month observations both pre- and post-2008 crisis. The mass in high trading frequency group G1 almost doubles in the post-crisis decade and stands at around 30% in 2017; the mass in low trading frequency group G1 shrunk accordingly. In the next chapters, we show that many bonds 'travel' across trading frequency groups during their lifetime in a non-intuitive way and have a puzzling relationship between changes in trading frequencies and prices.

[3]Our partition of Z into five intervals is immune to occasional distortions and short-lived jumps in trading activity. It would take roughly 4-5 additional business days with non-zero trading per month to take a particular bond to a higher trading frequency group.

3.3 (In)frequently traded bonds

3.3.1 Main characteristics

A widespread view links trading frequencies of corporate bonds with their maturity: right after issuance bonds trade actively on the secondary market, but after desired allocations are achieved trading activity slows down, and the closer the maturity is the less trading we observe. Such a pattern is indeed present in our data, but yet there is a large share of bonds whose trading activity evolves differently.

For every bond in our sample, we record a sequence of trading frequency groups (as defined in the previous chapter) that it belonged to. We are interested in the instances when bonds that presently trade rarely but were traded actively in the past start trading actively again. Table 3.2 counts the bonds that experienced this transition from frequent to infrequent trading, *and back*. There are about 3.9 thousand bonds in the entire sample, roughly 25% of all considered bonds, that make a trip from G1 (active trading) to G3-5 (inactive trading), and back to G1 at least once during the observed part of their lifetime. We call these bonds *(in)frequently traded bonds* or the (I)TBs. Roughly 2 out of 3 (I)TBs make the same trip, G1–G3-5–G1, at least twice in their life.

	Full sample	Pre-crisis	Post-crisis
.. 1 .. 5 .. 1 ..	808	72	441
.. 1 .. 4 .. 1 ..	2,150	399	1,515
.. 1 .. 4 or 5 .. 1 ..	2,172	404	1,529
.. 1 .. 3 or 4 or 5 .. 1 ..	3,886	1,058	2,985
.. 1 .. 3 or 4 or 5 .. 1 .. x2	2,487	481	1,861
Total no. of issues	14,287	8,348	11,462

Table 3.2. Number of bonds that travelled from a frequently traded category (G1) to infrequently traded categories (G3/4/5) and back (to G1). The sequences in rows indicate trip types. The first line is a trip from G1 to G5 and back to G1. The number of months spent in the intermediate states is unlimited. The second line is a trip from G1 to G4 and back to G1, etc. Columns represent bond subsamples. The last line is the total number of issues in the subsamples. The pre-crisis period is from Jan 2005 to Jun 2008; the post-crisis period is from Jan 2009 to Jun 2017.

Table 3.2 also shows how the fraction of bonds classified into the (I)TB subsample changes in pre- and post-crisis data treated separately. Throughout the chapter we define the pre-crisis period as Jan 2005 to Jun 2008 and the post-crisis period as Jan 2009 to Jun 2017.[4] In pre-crisis data, only 13% of bonds are the (I)TBs, in post-crisis data this fraction doubles. There is a substantially smaller fraction of the (I)TBs that make a G1–G3-5–G1 trip at least twice pre-crisis than post-crisis.

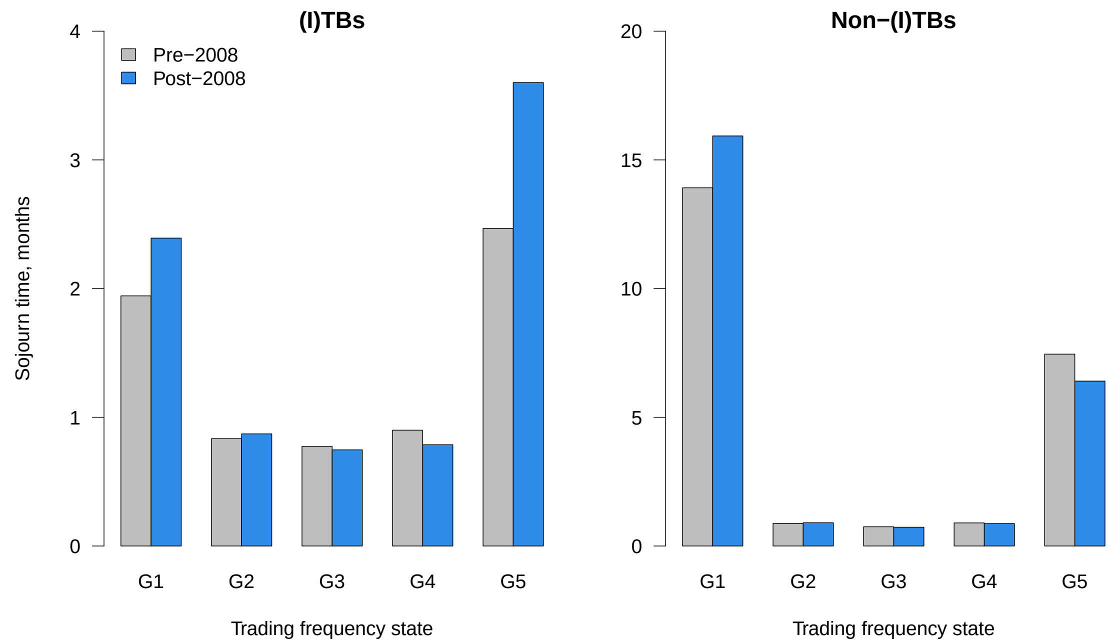


Figure 3.5. Number of months an average bond stays in a given trading frequency state (sojourn time). The underlying model is a five-state continuous time Markov chain with constant generator and instantaneous jumps to neighbouring states only.

To formally describe these 'waves' in trading activity we estimate a Markov model of the evolution of trading activity across five previously defined trading frequency states. The five-state Markov chain is defined in continuous time and instantaneous transitions are allowed to neighboring states only. Once we have the estimates of transition intensities, we compute monthly transition probabilities and average 'sojourn' times in each trading frequency states. Figure 3.5 presents the latter, Table C.3 in Appendix C.3 gives the former. Figure 3.5 shows that the (I)TBs stay in the active trading state G1 for about two months and in the inactive

[4] We remove the fall of 2008 from this data split to make sure that extreme crisis observations do not drive our results.

trading state G5 for three months on average. Both numbers are higher post-crisis than pre-crisis. The (I)TBs leave intermediate states within one month. The Non-(I)TBs stay in the boundary states for longer. For them, the average sojourn time is about fifteen months in the active trading state and twice less in the inactive trading state.

Surprisingly, we find very little difference between the (I)TBs and the Non-(I)TBs in primary bond characteristics. Table 3.3 presents means, medians, and inter-quartile range for the number of indicators. An average (I)TB bond in our sample has between 400 and 500 million USD outstanding amount, a credit rating between BBB+ and BBB (investment grade), it still has between 9 and 10 years to maturity (between 40% and 50% of its maturity at issuance has already elapsed), it is traded between 3 and 4 times a day (when traded at all) with an average volume per trade around 700 thousand USD, and in the same month we observe about 8 other outstanding bonds issued by the same firm. This description remains unchanged for an average Non-(I)TB bond except it has two more other outstanding bonds of the same issuer. There are more pronounced differences in the median outstanding amounts, relative age, and the number of trades between the (I)TBs and the Non-(I)TBs. The former tend to be 'younger', have higher outstanding amounts, and the number of trades per day. We experimented with different classification algorithms, including traditional logit regressions as well as boosted trees with random forests and more modern methods, to try to recover the classification into the (I)TBs and the Non-(I)TBs using primary bond characteristics only to conclude that it does not work. The information contained in sequences of trading frequency groups cannot be recovered from headline bond characteristics.

3.3.2 Trading volume and frequency

To give a better statistical description of the differences between the (I)TBs and the Non-(I)TBs we analyze in more details their trading records. Table 3.4 compares retail-size (trades $\leqslant$ 100'000 USD in volume) to institutional size (trades $>$ 100'000 USD in volume) trading volume in the (I)TBs and the Non-(I)TBs across trading frequency states. In all

	Mean		Median		IQR	
	(I)TB	Other	(I)TB	Other	(I)TB	Other
Amount outstanding, mln USD	465.4	411.5	400.0	200.0	350.0	358.1
Credit rating	8.7	8.1	8.0	8.0	4.0	3.0
Time since issuance, years	5.9	7.3	4.7	5.9	6.1	7.6
Time to maturity, years	9.5	9.8	6.2	6.1	9.6	10.8
Relative age, % of lifetime	43.6	48.2	40.8	47.6	45.5	49.2
Number of trades per business day	3.8	3.5	2.8	1.7	2.1	3.1
Average volume per trade, th USD	657.4	738.4	287.8	266.0	727.7	768.3
Number of bonds of the same issuer	9.2	11.2	7.0	6.0	9.0	10.0

Table 3.3. Descriptive statistics for (in)frequently traded bonds and all the other bonds. (In)frequently traded bonds are the bonds that made a trip G1–G3/4/5–G1. Credit rating is in conventional numerical score from 1 to 21: 1 corresponds to AAA, 8 to BBB+, 21 to C. IQR is the inter-quartile range.

states, both pre-crisis and post-crisis, aggregate monthly retail-size volume measured in % to institutional-size trading volume is substantially higher in the (I)TBs. The difference is the largest in the active trading state G1: here the average aggregate volume in small trades is roughly one-third of that in big trades for the (I)TBs and almost twice less in all other bonds.

	G1	G2	G3	G4	G5
	Pre-crisis				
(I)TB	31.43	24.05	19.58	12.77	6.36
Non-(I)TB	18.76	19.98	15.45	11.04	4.30
	Post-crisis				
(I)TB	31.69	23.84	20.77	17.46	9.27
Non-(I)TB	17.47	19.24	17.87	13.90	7.13

Table 3.4. Mean retail-size to institutional-size trading volume ratio, %. Institutional-size trades are above 100k USD. The sample is restricted to bond-month observations with positive institutional volume.

To link the extent of retail-size trading to changes in trading frequencies ΔZ_{it} (which leads to jumps between trading frequency states) we regress ΔZ_{it} on changes in trading volume split by size, direction, and counterparty. Using TRACE counterparty marker we classify every trade as either a buy transaction by a client from a dealer, or a sale by a client to a dealer, or an inter-dealer trade.[5] Each of the three categories is further split into two depending on the size of the trade.

	Dependent variable: $\Delta(Z_{it})$			
	(I)TB	Non-(I)TB	(I)TB	Non-(I)TB
	Pre-crisis		Post-crisis	
Δ(Client sell volume in big trades)$_{it}$	-0.24^{***}	-0.21^{***}	-0.24^{***}	-0.24^{***}
Δ(Client sell volume in small trades)$_{it}$	-11.14^{***}	-5.05^{***}	-11.60^{***}	-3.06^{***}
Δ(Client buy volume in big trades)$_{it}$	-0.28^{***}	-0.29^{***}	-0.48^{***}	-0.33^{***}
Δ(Client buy volume in small trades)$_{it}$	-13.09^{***}	-6.66^{***}	-7.86^{***}	-3.76^{***}
Δ(Inter-dealer volume in big trades)$_{it}$	-0.22^{***}	-0.19^{***}	-0.15^{***}	-0.14^{***}
Δ(Inter-dealer volume in small trades)$_{it}$	-6.23^{***}	-1.95^{***}	-7.42^{***}	-2.57^{***}
Month FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Observations	37,552	221,951	194,326	441,579
Adjusted R^2	0.14	0.11	0.11	0.08
Note:	*p<0.1; **p<0.05; ***p<0.01			
	Standard errors are clustered by the bond CUSIP.			

Table 3.5. Panel regressions of monthly changes in trading frequency ΔZ_{it} on trading volume split by size and type. Volumes are in % of the outstanding amount.

The sign of the relationship between changes in volume and ΔZ is straightforward: the bigger is the change in volume the more trading days we likely observe (hence, the lower ΔZ is). What matters more is how different this relationship is in the (I)TBs and the Non-(I)TBs. Table 3.5 shows that coefficients on trading volume in small trades are substantially higher in absolute value for the (I)TBs both pre-crisis and post-crisis. Z falls by 7 to 12 p.p. (percentage points) when an additional 1 p.p. of the bond outstanding amount is traded in small trades in a given month for an (I)TB compared to only 2 to 4 p.p. drop in Z for a Non-(I)TB. There is no such difference for big trades except for big trades that are client buy transactions. From Tables 3.4 and 3.5 we conclude that it takes more days to trade in small chunks the same volume of the (I)TBs than the Non-(I)TBs.

[5]TRACE contains trade reports by broker-dealers, hence every inter-dealer trade must appear twice in TRACE records. Only one such record remains in our sample after cleaning as in Dick-Nielsen (2014).

3.3.3 Mutual fund holdings, trading frequency, and illiquidity

We find differences between the (I)TBs and the Non-(I)TBs in mutual fund ownership and in the reaction of their ΔZ on changes in mutual fund holdings. Table 3.6 compares average mutual fund holdings of bonds in different trading frequency states (Table C.5 in Appendix C.3 presents additional descriptive statistics of mutual fund holdings). We find that mutual fund ownership ratios are higher for the (I)TBs both pre-crisis and post-crisis. The difference is especially pronounced in the least active trading state G5: on average 19% of the outstanding amount of a rarely traded (I)TB bond is held by mutual funds, 7 p.p. more than for an average Non-(I)TB bond.

	G1	G2	G3	G4	G5
Pre-crisis					
(I)TB	8.18	8.90	9.91	10.87	18.27
Non-(I)TB	7.17	9.44	9.22	8.63	9.81
Post-crisis					
(I)TB	12.08	12.29	12.63	12.92	19.11
Non-(I)TB	11.49	11.86	11.20	10.68	12.59

Table 3.6. Mean mutual fund holdings of bonds in different trading frequency states, % of the outstanding amount. Holdings are winsorized at 5% and 95%.

The difference between the (I)TBs and the Non-(I)TBs in the dispersion of fund ownership is even more striking. In Figure 3.6 we use a simple indicator: we count how many funds have non-zero holdings of a given bond in a given month depending on the trading frequency state. It turns out that for the (I)TBs this number is relatively constant across states. For instance, there are on average about 30 funds that own an (I)TB post-crisis (this number is close to 20 pre-crisis) regardless of whether the bond trades actively or not. This relationship is different for the Non-(I)TBs both pre- and post-crisis. Many more funds own a Non-(I)TB if it trades actively: there are more than 50 fund owners in G1 (more than for an (I)TB) compared to less than 20 in G5 (less than for an (I)TB).

If one assumes that the dispersion of mutual fund ownership is associated with information asymmetry in a given security (broader ownership arguably implies lower information asymmetry), than it should also be related to the autocovariance in returns which is a mea-

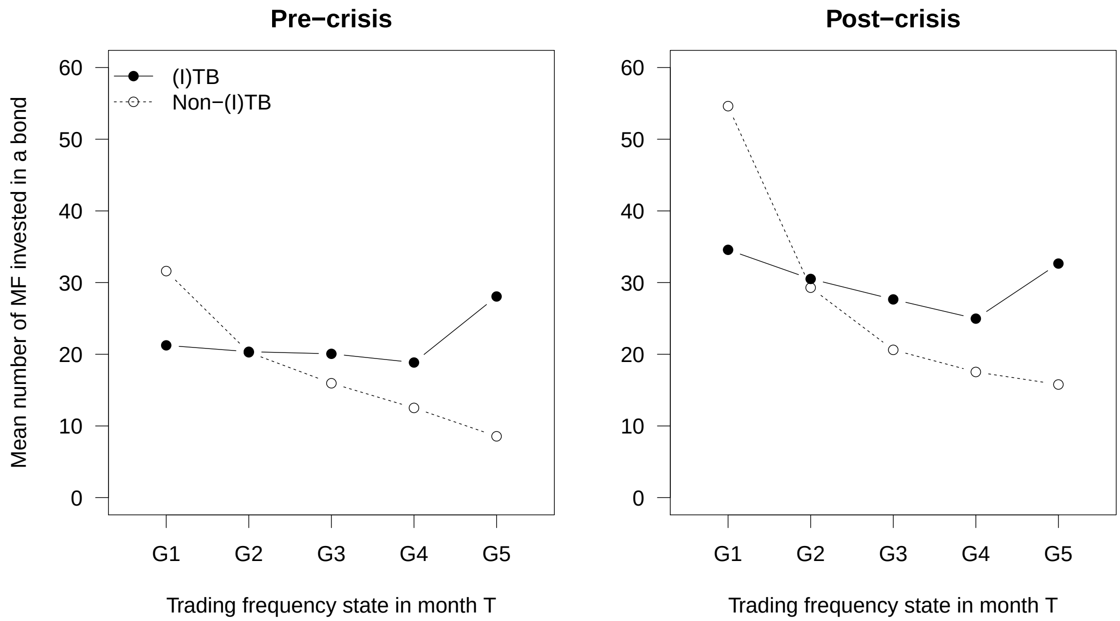


Figure 3.6. Mean number of mutual funds that hold the bond in different trading frequency states.

sure of illiquidity. For instance, Llorente et al. (2002) show that return autocovariance is more negative in stocks with higher information asymmetry (they are 'more illiquid'). In Table 3.7 we present the Bao et al. (2011) bond illiquidity measure (negative log return autocovariance), for bonds in our sample in different trading frequency states. For the Non-(I)TBs, illiquidity grows strongly with trading infrequency, which is in line with lower dispersion in fund ownership and higher information asymmetry in lower trading frequency states. For the (I)TBs, the dispersion of fund ownership is flatter across trading frequency states, so is the illiquidity. Post-crisis, the (I)TBs are more liquid than the Non-(I)TBs in states G3 and G4 according to the results in Table 3.7.

3.3.4 Changes in mutual fund demand and trading frequency

Changes in mutual fund net demand for corporate bonds significantly affect bond trading frequencies. Moreover, trading frequency tends to increase more when mutual funds are increasing their net demand for the (I)TBs rather than the Non-(I)TBs. Table 3.8 presents the regressions of changes in the trading frequency ΔZ_{it} of both types of bonds on changes

	G1	G2	G3	G4
Pre-crisis				
(I)TB	0.25	0.21	0.26	0.50
Non-(I)TB	0.14	0.24	0.27	0.37
Post-crisis				
(I)TB	0.26	0.27	0.36	0.50
Non-(I)TB	0.11	0.28	0.43	0.60

Table 3.7. Mean Bao et al. (2011) illiquidity measure for bonds in different trading frequency states. The illiquidity measure is a negative covariance of daily changes in volume-weighted average log-prices for months with at least 5 trading days. The illiquidity measure is winsorized at 0.1% and 99.9% in the entire sample. There are no observations in G5 because of the way the illiquidity measure is calculated.

in net purchases by mutual funds, insurance companies, and all other investors.[6] Among three types of investors considered, changes in net demand of mutual funds have the most substantial impact on changes in trading frequency. The effect is also stronger for the (I)TBs than for the Non-(I)TBs post-crisis. When mutual funds are buying 10 percentage points of the outstanding amount of a given bond more in a current month than in a previous month, Z falls by 2 percentage points for an (I)TB and by 1.4 percentage points for a Non-(I)TB.

	Dependent variable: $\Delta(Z_{it})$			
	(I)TB	Non-(I)TB	(I)TB	Non-(I)TB
	Pre-crisis		Post-crisis	
ΔMF net purchase$_{it}$	-0.10	-0.12^{**}	-0.20^{***}	-0.14^{***}
ΔIC net purchase$_{it}$	0.06	0.06^{***}	-0.004	0.01
ΔOther net purchase$_{it}$	-0.003^{**}	0.0005	-0.004^{**}	0.001
Month FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Observations	29,024	114,320	154,859	283,703
Adjusted R^2	0.02	0.004	0.02	0.01
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard errors are clustered by the bond CUSIP.			

Table 3.8. Panel regressions of monthly changes in trading frequency ΔZ_{it} on changes in net purchases by mutual funds, insurance companies, and other investors. Changes in net purchases are in % of the outstanding amount.

[6] We use interchangeably the terms 'net purchases' and 'net demand', both represent the difference between total buy and sell transactions. Mutual funds net demand is simply the change in total mutual fund holdings of a given bond. Net purchases by all other investors are the residual category. We know total net demand from TRACE, net mutual fund demand from processed SEC N-Q forms and net insurance companies demand from the NAIC data. Subtracting the last two from the first gives us net demand by investors other than U.S. mutual funds or insurance companies.

	Dependent variable: Δ(Net MF purchase)$_{it}$			
	(I)TB	Non-(I)TB	(I)TB	Non-(I)TB
	Pre-crisis		Post-crisis	
Δ(Client sell volume in big trades)$_{it}$	0.01	0.01***	0.01***	0.003
Δ(Client sell volume in small trades)$_{it}$	0.45	0.49**	0.20**	0.07
Δ(Client buy volume in big trades)$_{it}$	0.01***	0.01***	0.01***	0.01***
Δ(Client buy volume in small trades)$_{it}$	−0.03	−0.06	−0.04	−0.02
Δ(Inter-dealer volume in big trades)$_{it}$	0.01	0.01	−0.002	0.01*
Δ(Inter-dealer volume in small trades)$_{it}$	−0.07	−0.04	0.08	0.03
Month FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Observations	29,024	114,320	154,859	283,703
Adjusted R^2	0.05	0.04	0.02	0.02

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors are clustered by the bond CUSIP.

Table 3.9. Panel regressions of monthly changes in net mutual fund purchases on the trading volume split by size and type. Changes in net purchases and changes in volume are in % of the outstanding amount.

So far we have established that similar changes in small trading volume (especially in client sell trades) and in net mutual fund demand tend to have a stronger impact on changes in the trading frequency of the (I)TBs compared to the Non-(I)TBs. Now we ask, is there a relationship between changes in trading volume and net mutual fund demand at the first place? Table 3.9 regresses the latter on the former splitting volume by size and type as before. It turns out that post-crisis changes in net mutual fund demand are associated with changes in sell volume in small trades rather than any other type of volume, and more so for the (I)TBs. When *some clients* are selling 1 p.p. of the outstanding amount of an (I)TB more in a current month, we observe a significant increase in net mutual fund demand of 0.2 p.p. The effect is three times smaller and statistically insignificant for the Non-(I)TBs.

3.4 Trading frequency and public information about bond issuers and issues

There is a long list of potential issuer-level factors that might drive changes in bond trading frequencies. In this section, we first investigate how much variation in trading frequency changes is due to time-varying firm-level factors that we broadly refer to as 'corporate news' or simply 'news'. Corporate disclosures and public corporate events, media coverage, updates by equity analysts, spillovers from the equity market or the CDS market, etc. – any piece of information that is relevant for *all bonds of the same firm* we call the 'news'. Instead of trying to measure the news directly (which would be problematic given our broad definition of the news), we employ a modern econometric technique to select among time-varying firm-level dummies that proxy for the news and find that they explain only a small part of the variation in trading frequency changes. Then we demonstrate that the remaining *within-firm within-month* variation of bond trading frequencies is not well explained by bond-level characteristics either.

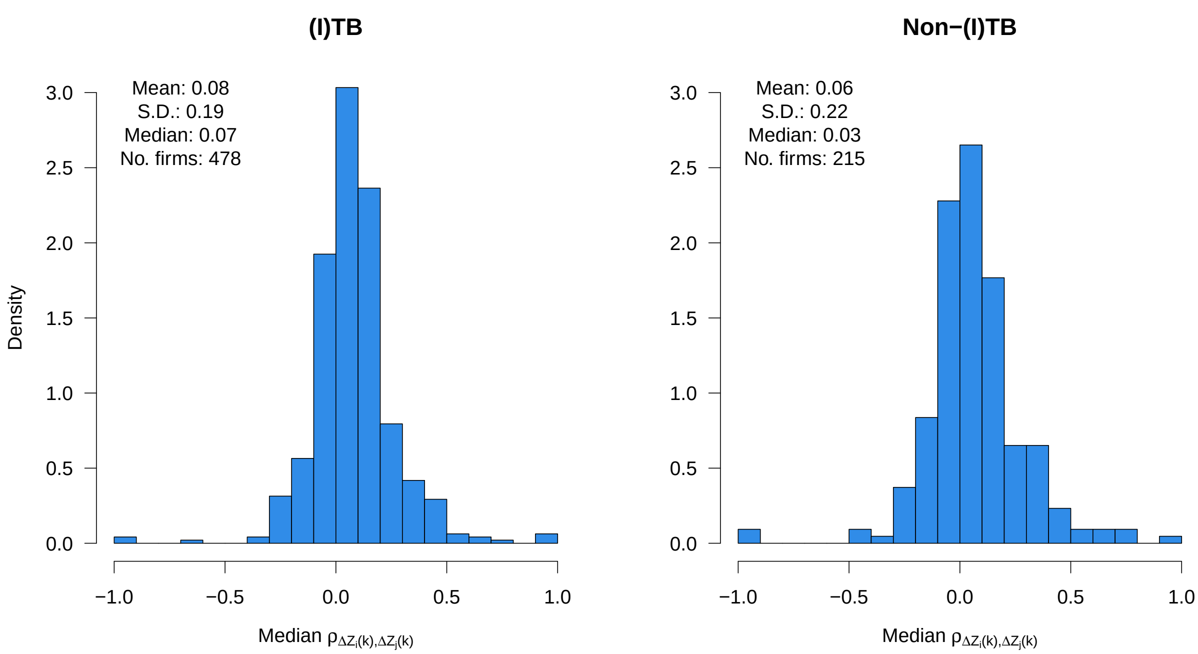


Figure 3.7. Cross-firm distribution of median pairwise correlation in ΔZ between different bonds of the same firm. We require at least 2 bonds of a given type of the same firm and at least 12 observation months per bond to compute correlations.

We start with a simple observation about correlations of changes in trading frequency $\Delta Z_{jt}(k)$ between bonds j of the same firm k. For all pairs of bonds of firm k we compute correlations $\rho_{\Delta Z_i(k),\Delta Z_j(k)}$ in trading frequency changes (we require at least 12 monthly observations per bond), then take the median pairwise correlation per firm $\rho(k)^{\text{median}}$, and plot the distribution of this number across firms on Figure 3.7. If changes in trading frequencies were mostly driven by firm-level factors, we would expect $\rho(k)^{\text{median}}$ to be positive and relatively high.[7] Instead, we observe on Figure 3.7 that the distributions are concentrated around zero with a small and insignificantly positive mean of 0.06–0.08 for both types of bonds considered. One should not expect high explanatory power of firm-level factors on changes in bond trading frequencies in such case.

To formally measure the explanatory power of corporate news and bond-level factors for changes in bond trading frequencies, we are using an econometric technique of double partialling-out introduced in Belloni, Chernozhukov, and Hansen (2014). Here is how we adapt it to our problem. Define ΔZ_i as the cross-sectional extract from $\{\Delta Z_{it}\}$ for any given month t. We will fit the models for ΔZ_i in the cross-section of bonds *independently for each month* t. The effect of firm-level news in month t will be captured by firm dummies D_f (where $f = 1, \ldots, F$; F being the total number of issuers) multiplied by respective coefficients γ_f to be estimated. The model for ΔZ_i at any given month is:

$$\Delta Z_i = (\beta_1 \Delta X_{1,i} + \cdots + \beta_P \Delta X_{P,i}) + (\gamma_1 D_1 + \cdots + \gamma_F D_F) + \epsilon_i,$$

where $\Delta X_1, \ldots, \Delta X_P$ are changes in bond-specific covariates of interest, and ϵ is orthogonal to both $\Delta X = (\Delta X_p)$ and $D = (D_f)$. There are *no restrictions* on the relationship among estimated coefficients in different months, the month-by-month cross-sectional estimations are fully independent from each other. The collection of estimates $\hat{\gamma}(t) = (\hat{\gamma}_1(t), \ldots, \hat{\gamma}_F(t))$ captures the total impact of time-varying firm-level factors on changes in bond trading frequencies. Our primary interest here is the joint explanatory power of firm dummies for ΔZ and the coefficients β, we have no interest in particular values of γ coefficients.

[7] We assume here that corporate news should affect different bonds of the same firm similarly, but do not test this assumption formally.

The cross-sectional model can be estimated with the OLS. But the OLS regression would suffer from over-fitting due to a relatively small number of bonds issued by each firm. We observe a median of 7 and 6 bonds per firm in the (I)TB and Non-(I)TB subsamples respectively. Firm dummies D_f would over-fit the data in the OLS regression, R^2 would be inflated and the estimates of β would be biased. To overcome the problem of too many explanatory variables relative to the sample size, Belloni et al. (2014) propose the following two-step procedure:

1. Project ΔZ and ΔX on D using some high-quality penalized regression procedure (we use LASSO here), compute the residuals $\Delta\tilde{Z} = \Delta Z - \Delta\hat{Z}$ and $\Delta\tilde{X} = \Delta X - \Delta\hat{X}$;
2. Run the OLS regression $\Delta\tilde{Z} = \Delta\tilde{X}\beta + u$, the estimate $\hat{\beta}_{OLS}$ is the consistent estimate of β of the original model.

In our case, the first stage projection of ΔZ on D is interesting per se. LASSO selects firms dummies and shrinks coefficients towards zero to avoid over-fitting. The intensity of shrinkage (LASSO penalty parameter) is chosen by 10-fold cross-validation. Each LASSO regression is run 30 times every month to explore the stability of the results. The explanatory power of this LASSO regression indicates what portion of the variation of ΔZ is due to corporate news. The second stage OLS regression of $\Delta\tilde{Z}$ on $\Delta\tilde{X}$ investigates how the residuals of bond-level covariates unexplained by firm dummies affect the residuals of changes in trading frequencies.

Figure 3.8 presents the R^2 from the first stage Ridge regressions of ΔZ on D for (in)frequently traded bonds (the results for the Non-(I)TBs are similar). The R^2 varies over time from 0 to about 20% (shaded area), with the smoothed median value being close to 5% before 2010 and even lower after that. It means that the impact of time-varying firm characteristics on the frequency of bond trading is very limited. Even if we run a plain OLS regression of ΔZ on D on the first-stage, the over-fitted R^2 is around 40% in the post-crisis period (dashed-dotted line on Figure 3.8). Observe also that the adjusted R^2 of the OLS regression (dotted line) is of the same order of magnitude as the R^2 from the LASSO first-stage regression. Hence,

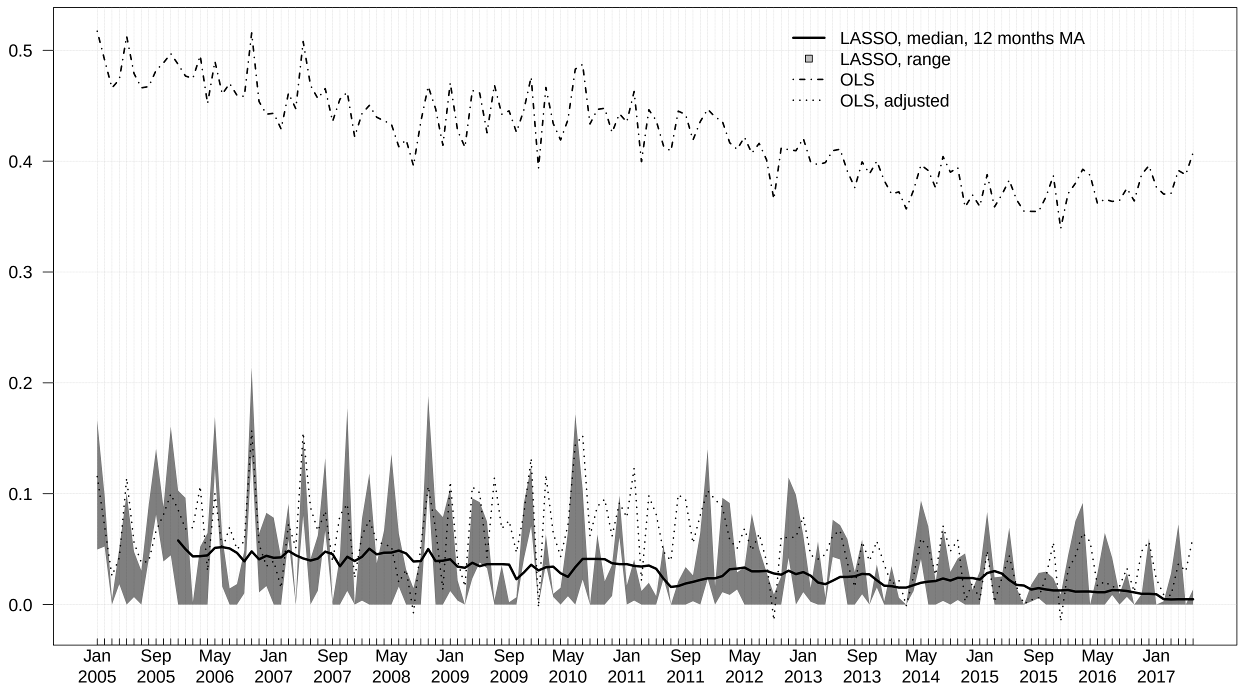


Figure 3.8. R^2 of the first stage cross-section regressions of ΔZ_i on firm dummies. The penalty parameter for the first stage LASSO regression is chosen by 10-fold cross-validation. Each regression is estimated 30 times to investigate the stability of the results. The range of R^2 generated by these 30 runs is the shaded grey area on this plot. The solid black line is the median value of that range after smoothing with 12-month backwards-looking moving average. Dashed and dashed-dotted lines are respectively adjusted R^2 and simple R^2 from (over-fitted) cross-section OLS-regressions. The sample is the (I)TBs.

changes in trading frequency of the (I)TBs remain largely unexplained by corporate news, broadly defined.

The second stage regression of $\Delta\tilde{Z}$ on $\Delta\tilde{X}$ is presented on Figure 3.9. We consider three explanatory variables: changes in outstanding amount (size), credit rating, and relative age (% of bond lifetime that has already passed at the measurement date). These variables were pre-selected by running multiple panel models of ΔZ on bond- and firm-level covariates with independent firm and time fixed effects; they turned out to be the most significant ones across different specifications. Solid lines on Figure 3.9 present point estimates of the coefficients on corresponding covariates. The signs of coefficients on Figure 3.9 are similar to a simple panel model with independent firm and time fixed effects (see Table C.6 in Appendix C.3): bond redemptions and bond ageing are associated with lower trading frequencies. Yet, Figure 3.9

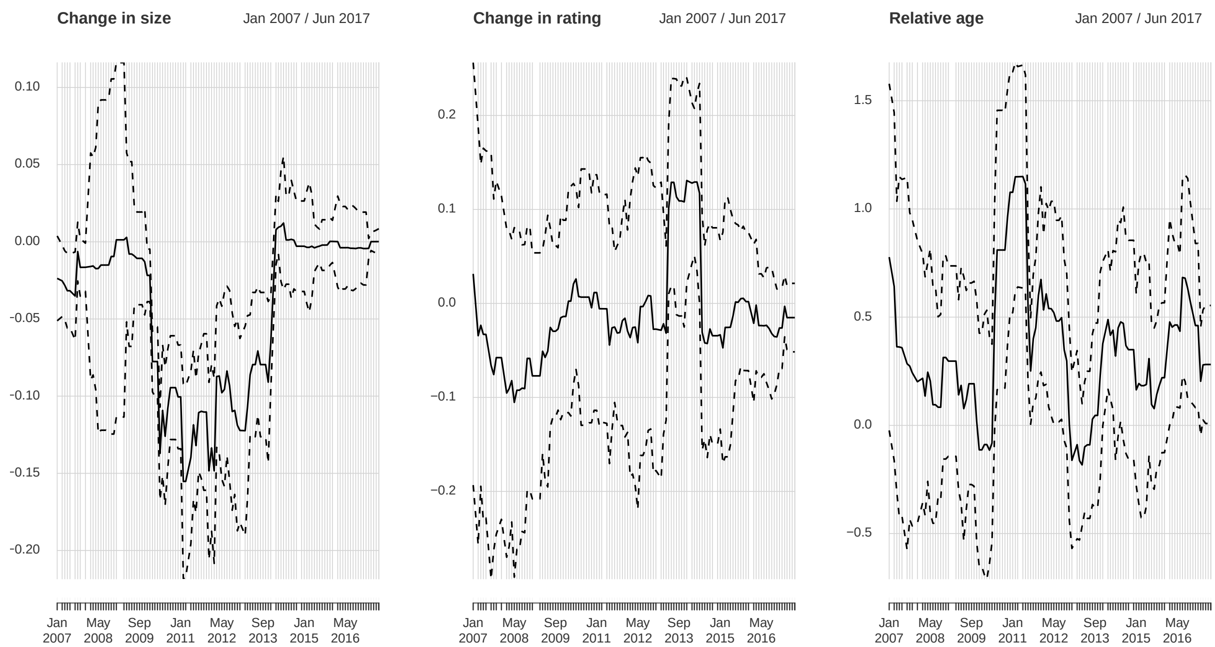


Figure 3.9. Coefficients on candidate covariates in cross-section second stage OLS regressions of $\Delta\tilde{Z}$. Solid lines are 12-month moving-average point estimates. Dashed lines are 12-month moving averages of 2 standard error bounds around point estimates. Some months have no variability in covariates; they are excluded from estimation. Change in size is the % change in the outstanding amount month-on-month. Age is the time elapsed since issuance as a fraction of total maturity at issuance. Credit rating is on the conventional numerical scale from 1 (AAA) to 21 (D), a unit change represents an upgrade or a downgrade by one notch. The sample is the (I)TBs.

says that these effects are not stable over time and the statistical significance is often absent. The effect of credit rating changes on trading frequencies is small and nowhere significant (unlike in the panel model with independent firm and time fixed effects). The explanatory power of $\Delta\tilde{X}$ for $\Delta\tilde{Z}$ in the second-stage regressions is small either: mean R^2 is close to 5%. So, the evidence presented on Figures 3.7–3.9 suggests that corporate news and major bond characteristics can explain only a small portion of changes in bond trading frequencies ΔZ.

We further explore within-firm within-month variation in bond trading frequencies in the subsample of issuers with many outstanding bonds (over-fitting with firm-dummies is less likely here) to confirm our previous findings. We require at least 10 bonds per firm to be observed for at least 12 months each for the issuer to be included in the sample. There are

50 and 150 firms that satisfy these criteria for the (I)TBs and non-(I)TBs respectively. For each of these firms k we run *separately* a fixed-effects panel model:

$$\Delta Z_{jt}(k) = (\beta_1 \Delta X_{1,jt}(k) + \cdots + \beta_P \Delta X_{P,jt}(k)) + (\tau_1 D_1 + \cdots + \tau_T D_T) + u_{jt},$$

where $\Delta Z_{jt}(k)$ is the change in trading frequency of bond j of firm k in month t, ΔX are bond-specific factors, and month dummies D capture time fixed effects for all bonds of firm k. Under the assumption that news affect changes in trading frequencies of all bonds of the same firm similarly, time fixed effects in the regression above capture the effect of corporate news on firm's k bond trading frequencies.

	Mean	Med.	Min.	Max.	No. firms
	R^2				
(I)TB	0.15	0.14	0.09	0.30	50
Non-(I)TB	0.20	0.17	0.04	0.95	150
	Adjusted R^2				
(I)TB	0.06	0.05	-0.06	0.20	50
Non-(I)TB	0.06	0.05	-0.43	0.77	150

Table 3.10. Explanatory power of firm-level regressions for ΔZ_{it} in the subsample of issuers with many outstanding bonds. The estimated model includes changes in amount outstanding and credit rating, relative age, changes in the number of outstanding bonds of the same issuer, coupon dummy, and month fixed effects as explanatory variables. The model is estimated separately for each firm, hence, the dataset in each estimation consists of different bonds of the same firm observed in different months. We require at least 10 bonds of certain type to be observed in at least 12 months for a firm to be included in the sample. The number of firms in the last column shows how many firms satisfy these criteria for two types of bonds considered.

We estimate the firm-level models with the OLS. The OLS still over-fits the data, but now we have around 10 observations per estimated coefficient (if the panel is balanced). Table 3.10 presents R^2 and adjusted R^2 from firm-level regressions. These numbers show the percentage of variation in changes in trading frequency that is explained by corporate news and bond-specific factors *combined*. The average R^2 in Table 3.10 is around 15% for (in)frequently traded bonds and 20% for all other bonds. Median R^2 is a bit lower than the mean, adjusted R^2 are around 5-6% on average. These adjusted R^2 values are in line with the evidence provided earlier in this chapter. About 5% – this is how much variance of changes in bond trading frequency we can credibly explain with corporate news and bond-specific factors. We believe that this number is quite low, and conclude that changes in bond trading frequency are mostly due to factors unrelated to bond- or firm-level characteristics and corporate news. Hence, spikes and dry-ups in bond trading activity are probably more related to *who* trades the bonds rather than to *what* bonds are traded.

3.5 Trading frequency and returns

This chapter describes a puzzling observation: when the (I)TBs jump to states with more (less) frequent trading, they generate positive (negative) returns that are not explained by institutional trading flows and exposure to risk factors. There is no such effects for the Non-(I)TBs.

We have already presented the phenomenon briefly in the introduction. Figure 3.2 shows that the effect is two-fold. The states with more active trading in month $T-1$ are associated with higher returns in month T, and these returns are higher or lower if trading frequency increases or decreases in month T. Table C.7 presents the same result in a more elaborate form comparing returns across 25 possible combinations of trading frequency states in months $T-1$ and T. In the rest of the chapter, we demonstrate that the effect appears only after the 2008 crisis and is not subsumed by the exposure of the (I)TBs to main risk factors and institutional flows.

We start with Figure 3.10, the analog of Figure 3.2, where instead of mean returns we plot bond alphas. Here we first compute value-weighted return time series for 25 bond baskets based on all combinations of five trading frequency states in the previous and in the current month.[8] We regress each of these 25 time series on 4 time series of Bai et al. (2019)

[8]These bond baskets are not investable since the trading frequency state in month T is not know apriori. Switching from value-weighted to equally-weighted returns does not change the results.

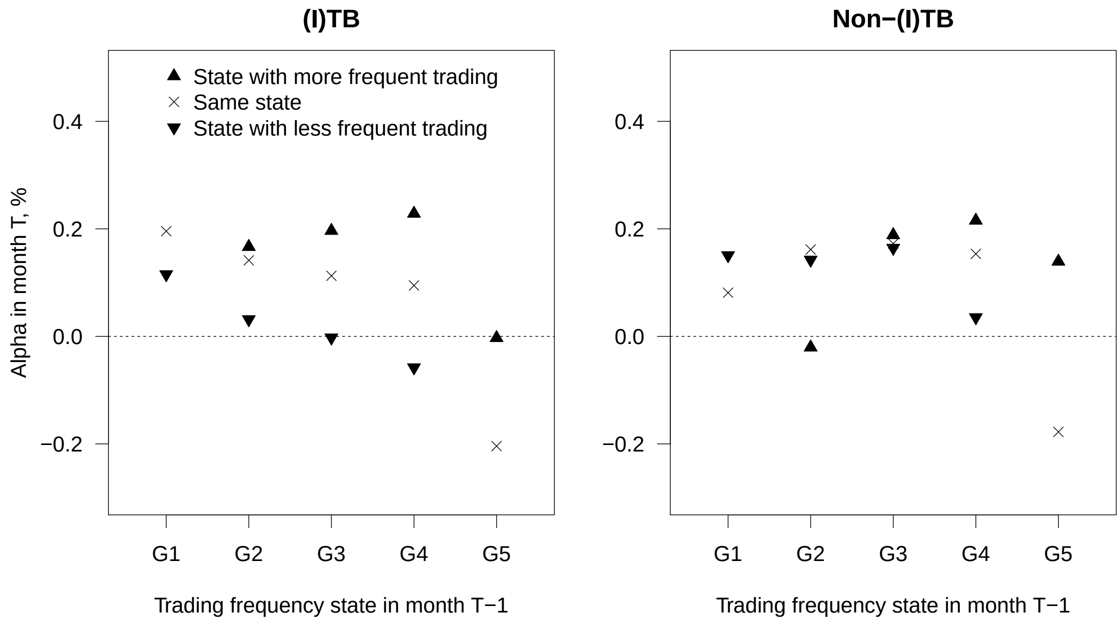


Figure 3.10. Estimated alphas and trading frequency jumps. The underlying model is the Bai et al. (2019) model. Bond baskets here are not investable since the trading frequency state in month T is not known apriori. Returns are computed by weighting individual bond returns in excess of the 3-month T-Bill rate by the market value of issues.

pricing factors and extract alphas.[9] All 25 estimated alphas are presented in Table C.8 in Appendix C.3, Figure 3.10 presents the same result in a more intuitive form. We observe the same relationship between alphas and trading frequency states as before and, again, only for the (I)TBs. States with higher trading frequency in month $T-1$ are associated with higher abnormal returns in month T, and if an (I)TB jumps to a higher trading frequency state in month T this abnormal return is even higher (lower if trading frequency falls).

It turns out that the relationship between trading frequency jumps and returns appears only after the 2008 crisis. To demonstrate it formally we run panel regressions for excess returns pre- and post-crisis separately. Our basic regression has the following form:

$$R_{it} = \sum_{s=2}^{5} \left(\beta_s \cdot D_{i,t-1}^{\text{State = s}}\right) + \sum_{s=1}^{5} \left(\gamma_s \cdot D_{i,t-1}^{\text{State = s}} \cdot \text{Jump}_{it}\right) + \text{Month FE} + \text{Issuer FE} + \epsilon_{it}.$$

[9]These are the market, default, liquidity, and credit factors. The last 3 are constructed by double sorting on 36-month 5% VaR, Bao et al. (2011) illiquidity measure, and credit rating in different combinations.

	Dependent variable: R_{it}			
	(I)TB	Non-(I)TB	(I)TB	Non-(I)TB
	Pre crisis		Post crisis	
$\text{State}_{i,t-1} = \text{G2}$	−0.050	−0.103**	−0.106***	−0.058*
$\text{State}_{i,t-1} = \text{G3}$	−0.230***	−0.245***	−0.151***	−0.032
$\text{State}_{i,t-1} = \text{G4}$	−0.249**	−0.301***	−0.271***	−0.035
$\text{State}_{i,t-1} = \text{G5}$	−0.517	−0.482***	−0.651***	−0.106*
$(\text{State}_{i,t-1} = \text{G1}) \times \text{Jump}_{it}$	0.160***	0.084	0.148***	0.002
$(\text{State}_{i,t-1} = \text{G2}) \times \text{Jump}_{it}$	0.066*	0.012	0.077***	0.056
$(\text{State}_{i,t-1} = \text{G3}) \times \text{Jump}_{it}$	0.115**	0.090***	0.077***	0.088**
$(\text{State}_{i,t-1} = \text{G4}) \times \text{Jump}_{it}$	0.108	0.153***	0.172***	0.060**
$(\text{State}_{i,t-1} = \text{G5}) \times \text{Jump}_{it}$	0.264	0.120**	0.479***	0.109**
Month FE	YES	YES	YES	YES
Issuer FE	YES	YES	YES	YES
Observations	22,446	54,548	113,598	148,280
Adjusted R^2	0.097	0.122	0.191	0.205
Note:	*p<0.1; **p<0.05; ***p<0.01 SEs are clustered by bond CUSIP.			

Table 3.11. Regressions of excess returns on trading frequency characteristics. 'Jump' is the integer variable that equals the difference in trading frequency state numbers in months $t-1$ and t. For instance, if the bond jumps from state G3 to state G1, the $\text{Jump}_{it} = 2$. The reverse jump has the value of -2.

Here $D_{i,t-1}^{\text{State = s}}$ is a dummy variable that takes the value of 1 if the bond i is in the trading frequency state $s \in 1, 2, \ldots, 5$ in month $t-1$. Coefficients β_s capture the relationship between past trading frequency states and current excess returns relative to excess returns in the most active trading state G1. Jump_{it} is the integer variable that equals the difference in trading frequency state numbers in months $t-1$ and t. For instance, if the bond jumps from state G3 to state G1, the $\text{Jump}_{it} = 2$. The reverse jump has the value of -2. If the bond stays in the same trading frequency state then $\text{Jump}_{it} = 0$. Hence, coefficients γ_s capture additional returns associated with trading frequency jumps in month t relative to returns in the case when trading frequency state does not change.[10]

Table 3.11 shows that only for the (I)TBs and only post-crisis all $\hat{\beta}_s$ and $\hat{\gamma}_s$ are highly significant. Coefficients $\hat{\beta}_s$ monotonically decrease with s from -10 b.p. to -65 b.p. For instance, average returns in month t of the bonds that were in state G2 in month $t-1$ are 10 b.p. lower than of the bonds that were in state G1 in month $t-1$. Coefficients $\hat{\gamma}_s$ are positive suggesting that that jumps towards more active trading yield additional positive

[10] We could also tell the same story with 25 estimated dummies for all the combinations of trading frequency states in months $t-1$ and t. We prefer this form with jump variables for its concise presentation.

returns and jumps towards less active trading result in lower returns. The absolute value of these additional returns is around 12 b.p. for jumps from states G1–4. Observe that pre-crisis the effect is less significant or absent.[11]

	Dependent variable: R_{it}			
	(I)TB	Non-(I)TB	(I)TB	Non-(I)TB
	Before Jun 2008		After Jan 2009	
ΔNet purchase in big trades$_{i,t}$	0.005	0.004	0.012***	0.008**
ΔNet purchase in small trades$_{i,t}$	−0.824***	−0.259	−0.240***	−0.159**
State$_{i,t-1}$ = G2	−0.038	−0.100**	−0.102***	−0.057
State$_{i,t-1}$ = G3	−0.223***	−0.243***	−0.147***	−0.031
State$_{i,t-1}$ = G4	−0.257**	−0.299***	−0.270***	−0.035
State$_{i,t-1}$ = G5	−0.579*	−0.485***	−0.669***	−0.108*
(State$_{i,t-1}$ = G1) × Jump$_{it}$	0.176***	0.092	0.152***	0.003
(State$_{i,t-1}$ = G2) × Jump$_{it}$	0.084**	0.018	0.081***	0.058
(State$_{i,t-1}$ = G3) × Jump$_{it}$	0.137***	0.095***	0.082***	0.090***
(State$_{i,t-1}$ = G4) × Jump$_{it}$	0.131	0.157***	0.179***	0.063**
(State$_{i,t-1}$ = G5) × Jump$_{it}$	0.338*	0.128**	0.501***	0.114**
Month FE	YES	YES	YES	YES
Issuer FE	YES	YES	YES	YES
Observations	22,446	54,548	113,598	148,280
Adjusted R^2	0.100	0.122	0.191	0.206
Note:	*p<0.1; **p<0.05; ***p<0.01 SEs are clustered by bond CUSIP.			

Table 3.12. Regressions of returns on changes in net purchases grouped by size and trading frequency characteristics. Trades with less than 100k volume are small trades. 'Jump' is the integer variable that equals the difference in trading frequency state numbers in months $t-1$ and t. For instance, if the bond jumps from state G3 to state G1, the Jump$_{it} = 2$. The reverse jump has the value of -2.

In Tables 3.12 and 3.13 we add either changes in net buy volume in big and small trades or changes in net demand by institutional investors to the baseline regression specification. Changes in volume in small trades and changes in net mutual fund purchases are related to jumps in trading frequency. Hence, one might expect them to explain some of the effects of trading frequency jumps on returns. It does not happen, at least for the (I)TBs post-crisis. Remarkable though that the signs at changes in net buy volume in big and small trades in Table 3.12 are opposite. Increases in net buy volume in big trades are associated with higher returns while increases in net buy volume in small trades are associated with lower returns (equivalently, increases in net sell volume in small trades occur in bond-months with higher

[11] Figure C.1 in Appendix C.3 plots cumulative returns on some of the 25 bond baskets described above. These graphs also demonstrate that returns associated with trading frequency jumps accrue only after 2008.

	Dependent variable: R_{it}			
	(I)TB	Non-(I)TB	(I)TB	Non-(I)TB
	Before Jun 2008		After Jan 2009	
ΔMF net purchase$_{i,t}$	0.030**	0.015	0.001	0.014
ΔIC net purchase$_{i,t}$	0.007	0.003	0.008	0.011*
ΔOther net purchase$_{i,t}$	−0.001	0.0003	−0.0005	0.00003
State$_{i,t-1}$ = G2	−0.061	−0.152***	−0.105***	−0.045
State$_{i,t-1}$ = G3	−0.246***	−0.297***	−0.152***	0.00001
State$_{i,t-1}$ = G4	−0.181	−0.351***	−0.289***	0.019
State$_{i,t-1}$ = G5	−0.747	−0.621***	−0.846***	0.048
(State$_{i,t-1}$ = G1) × Jump$_{it}$	0.175***	0.102	0.147***	0.025
(State$_{i,t-1}$ = G2) × Jump$_{it}$	0.059	0.041	0.075***	0.057
(State$_{i,t-1}$ = G3) × Jump$_{it}$	0.140**	0.111***	0.090***	0.082**
(State$_{i,t-1}$ = G4) × Jump$_{it}$	0.023	0.161***	0.227***	0.064*
(State$_{i,t-1}$ = G5) × Jump$_{it}$	0.552	0.122*	0.656***	0.054
Month FE	YES	YES	YES	YES
Issuer FE	YES	YES	YES	YES
Observations	19,221	44,182	98,324	131,236
Adjusted R^2	0.100	0.122	0.191	0.204
Note:	*p<0.1; **p<0.05; ***p<0.01 SEs are clustered by bond CUSIP.			

Table 3.13. Regressions of returns on changes in net institutional demand and trading frequency characteristics. 'MF' stand for mutual funds, 'IC' for insurance companies. 'Jump' is the integer variable that equals the difference in trading frequency state numbers in months $t-1$ and t. For instance, if the bond jumps from state G3 to state G1, the $\text{Jump}_{it} = 2$. The reverse jump has the value of -2.

returns). A rise of net sales in small trades of 1 p.p. of the outstanding amount translates into 25 b.p. of excess return in the (I)TBs and 16 b.p. in the Non-(I)TBs.

We like the following explanation for the positive impact of changes in big net buys and small net sells on returns, especially in ITBs. We have established before that when mutual funds increase their net purchases of the (I)TBs (arguably for non-informational reasons), some other agents sell more of these bonds in small trades. Now we know that it also pushes prices up. Given the time frame we are looking at, we tend to think that increases in sell volumes in small trades represent profit-taking by hedge funds that were entering the corporate bond market actively in 2008 and 2009 and closing positions several years later. The fact that trading frequency jumps still affect returns even when we control for changes in volumes and institutional flows suggests that there was potentially some friction in the reallocation of bonds from hedge funds to mutual funds that pushed prices even further up. It might be lower dealer inventory levels and longer intermediation chains post-crisis.

	Dependent variable: R_{it}	
	(I)TB	Non-(I)TB
$\text{VaR}_{i,t-1}$	0.084***	0.098***
$\text{Rating}_{i,t-1}$	0.204***	0.128***
$\text{Illiquidity}_{i,t-1}$	0.090**	0.078
$\text{State}_{i,t-1} = \text{G2}$	−0.071*	−0.007
$\text{State}_{i,t-1} = \text{G3}$	−0.139***	−0.061
$\text{State}_{i,t-1} = \text{G4}$	−0.297**	−0.156
$\text{State}_{i,t-1} = \text{G5}$	0.167***	0.090
$(\text{State}_{i,t-1} = \text{G1}) \times \text{Jump}_{it}$	0.107***	0.128
$(\text{State}_{i,t-1} = \text{G2}) \times \text{Jump}_{it}$	0.100**	0.163**
$(\text{State}_{i,t-1} = \text{G3}) \times \text{Jump}_{it}$	0.061	0.062
Month FE	YES	YES
Issuer FE	YES	YES
Observations	42,792	47,441
Adjusted R^2	0.197	0.201
Note:	*p<0.1; **p<0.05; ***p<0.01 SEs are clustered by bond CUSIP.	

Table 3.14. Regressions of post-crisis returns on riskiness and trading frequency characteristics. VaR is the 36-month rolling 5% value at risk (second smallest return). Rating is the numerical score from 1 to 21. Illiquidity is the Bao et al. (2011) measure. 'Jump' is the integer variable that equals the difference in trading frequency state numbers in months $t-1$ and t. For instance, if the bond jumps from state G3 to state G1, the $\text{Jump}_{it} = 2$. The reverse jump has the value of -2.

Indirectly supporting this point of view, we present in Table 3.14 the baseline regression for returns extended with corporate bond risk proxies from Bai et al. (2019): bond-level Value at Risk, Bao et al. (2011) illiquidity, and credit rating, all lagged one period. If the effect of trading frequency jumps on returns was due to the exposure of the (I)TBs to these risk factors rather than to a complicated interplay of institutional liquidity trading and low inventory issues, then we would not observe significant coefficients at trading frequency levels and jumps in the extended regression. Table 3.14 shows that it happens only to a minimal extent. For the (I)TBs, all $\hat{\beta}_s$ and $\hat{\gamma}_s$ except for one are still significant.

3.6 Concluding Remarks

In this chapter, we analyzed a large subset of plain-vanilla fixed coupon corporate bonds that experience prolonged swings in trading activity long after issuance. We called these bonds that 'travel' from active to inactive trading *and back* (in)frequently traded, or the

(I)TBs, and attempted to describe statistically the dimensions along which the (I)TBs are different from the Non-(I)TBs. It turned out that headline bond characteristics like size, maturity, and credit rating are not much different in our two subsamples of bonds. We found substantial differences between the (I)TBs and the Non-(I)TBs in the structure of their trade flow and institutional ownership.

First, we demonstrated that in the (I)TBs higher volumes are traded via small trades, and it takes more days to trade the same additional volume in small trades in the (I)TBs than in the Non-(I)TBs. The latter might indicate that intermediation chains are longer in the (I)TBs. Second, we found that the (I)TBs are more likely to be owned by mutual funds. Remarkably, there's a relatively constant number of funds that hold an (I)TB in any trading frequency state, unlike a Non-(I)TB that is held by substantially fewer funds when it trades infrequently. Related to that, the illiquidity of the (I)TBs grows very moderately with trading infrequency compared to the Non-(I)TBs. Third, we showed that positive changes in mutual fund net demand are associated with positive changes in sell volume in small trades and more frequent trading. Next, we argued that time-varying firm-level and bond-level characteristics were able to explain only a minor fraction of variation of changes in trading frequency, and so the long-lasting waves of trading activity we documented were not attributed to public news about the issuers or the issues.

Finally, we documented that the (I)TBs yield abnormal returns that relate to the swings of trading activity in a way unexplained by common bond-risk factors and institutional flows. When the (I)TBs jumped to states with more (less) frequent trading, they generated positive (negative) returns in the after-crisis period. There were no such effects for the Non-(I)TBs.

Overall, it seems that the (I)TBs happened to be the bonds that were in high demand among mutual funds, especially in the post-crisis period. We tend to think that the sell volume in small trades that goes up together with the (I)TBs trading frequency suggests that mutual funds were ultimately purchasing these bonds from smaller investors like hedge funds that populated the market in the aftermath of the 2008 crisis. Given dealers' preferences for low inventory levels after 2008, the intermediation between selling small investors and buying mutual funds was relatively slow and contributed to abnormal returns of the (I)TBs.

Bibliography

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

Alexander, Gordon J, Amy K Edwards, and Michael G Ferri, 2000, The determinants of trading volume of high-yield corporate bonds, *Journal of Financial Markets* 3, 177–204.

Aruoba, S. Borağan, Francis X. Diebold, and Chiara Scotti, 2009, Real-time measurement of business conditions, *Journal of Business & Economic Statistics* 27, 417–427.

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

Asquith, Paul, Thom Covert, and Parag Pathak, 2013b, The effects of mandatory transparency in financial market design: Evidence from the corporate bond market, Working Paper 19417, National Bureau of Economic Research.

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.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, 2016, Measuring economic policy uncertainty, *The Quarterly Journal of Economics* 131, 1593.

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.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, 2014, Inference on treatment effects after selection among high-dimensional controls, *The Review of Economic Studies* 81, 608–650.

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.

Bessembinder, Hendrik, William Maxwell, and Kumar Venkataraman, 2006, Market transparency, liquidity externalities, and institutional trading costs in corporate bonds, *Journal of Financial Economics* 82, 251–288.

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.

Collin-Dufresne, Pierre, Robert S. Goldstein, and J. Spencer Martin, 2001, The determinants of credit spread changes, *The Journal of Finance* 56, 2177–2207.

d'Avernas, Adrien, 2017, Disentangling credit spreads and equity volatility, Working paper, available at http://www.adriendavernas.com/papers/adriendavernas_jmp.pdf.

De Santis, Roberto A., 2017, Relative excess bond premium and economic activity, Working paper, an earlier version available at https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1930.en.pdf?6fa2032a4579de4ba73e9f40bfe24111.

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, Peter Feldhütter, and David Lando, 2012, Corporate bond liquidity before and after the onset of the subprime crisis, *Journal of Financial Economics* 103, 471 – 492.

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

Diebold, Francis X, and Roberto S Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business & Economic Statistics* 13, 253–263.

Duan, Jin-Chuan, and Tao Wang, 2012, Measuring distance-to-default for financial and non-financial firms, *Global Credit Review* 02, 95–108.

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, 2007a, Corporate bond market transaction costs and transparency, *The Journal of Finance* 62, 1421–1451.

Edwards, Amy K, Lawrence E Harris, and Michael S Piwowar, 2007b, 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* .

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2007, The U.S. treasury yield curve: 1961 to the present, *Journal of Monetary Economics* 54, 2291 – 2304.

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.

Harris, Lawrence, 2015, Transaction costs, trade throughs, and riskless principal trading in corporate bond markets, Working paper, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2661801.

Hendershott, Terrence, Roman Kozhan, and Vikas Raman, 2019, Short selling and price discovery in corporate bonds, *Journal of Financial and Quantitative Analysis* 1–39.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *The Review of Financial Studies* 5, 357.

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.

Mariano, Roberto S., and Yasutomo Murasawa, 2003, A new coincident index of business cycles based on monthly and quarterly series, *Journal of Applied Econometrics* 18, 427–443.

Merton, Robert C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *The Journal of Finance* 29, 449–470.

Newey, Whitney, and Kenneth West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–08.

Nozawa, Yoshio, 2017, What drives the cross-section of credit spreads?: A variance decomposition approach, *The Journal of Finance* 72, 2045–2072.

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

Appendix of Chapter 1

A.1 Aspects of the model

A.1.1 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 $t + 1$.[1] 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

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

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

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

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

Then the expression for returns becomes:

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

which is equation (1.4). I set $\bar{p}=0$ (the steady-state bond price is par), then $\theta = \frac{1}{1+C}$.

A.1.2 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

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

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 $f_{g|\tilde{p},s}$:

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

Expanding the square and collecting terms, one gets:

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

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:

$$\mathbb{E}_t^{(2)}\left[g_t|\tilde{p}_t, s_t\right] = \underbrace{\frac{\sigma_g^2}{\sigma_g^2 + b^2\sigma_z^2}}_{\equiv\gamma}(g_t + bz_t),$$

$$\mathbb{V}_t^{(2)}\left[g_t|\tilde{p}_t, s_t\right] = (1-\gamma)\sigma_g^2.$$

A.1.3 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 $\mathbb{E}_t^{(1)}\left[r_{t+1}\right]$ and 0, and variances $\mathbb{V}_t^{(1)}\left[r_{t+1}\right]$ 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

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

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

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

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

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

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

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

A.1.4 Existence of the equilibrium

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

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

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:

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

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:

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

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

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

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

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

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 1.6 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 $\frac{d}{da}\left[e^*(a)\right] > 0$, i.e.:

$$\begin{aligned}
v' - v'k - vk' > 0 &\Leftrightarrow \\
v'(1-k) > vk' &\Leftrightarrow \\
\frac{v'}{v} > \frac{k'}{1-k} &\Leftrightarrow \\
\frac{v'}{v} > \frac{k'(1+k)}{1-k^2} &\Leftrightarrow \\
\frac{v'}{v} > \frac{-\frac{1}{2k}B^2\psi'(1+k)}{B^2\psi} &\Leftrightarrow \\
\frac{v'}{v} > -\frac{1}{2}\frac{\psi'\left(1+\frac{1}{k}\right)}{\psi} &\Leftrightarrow \\
-\frac{2a-\gamma}{a(a-\gamma)} > -\frac{1}{2}\frac{\psi'\left(1+\frac{1}{k}\right)}{\psi} &\Leftrightarrow \\
-\frac{2a-\gamma}{a(a-\gamma)} > -\frac{\frac{\sigma_f^2}{\sigma_s^2}(a-\gamma) + \frac{\sigma_g^2+b^2\sigma_z^2}{\sigma_s^2}a(a-\gamma)(2a-\gamma)}{(a-\gamma)^2\left(\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2+b^2\sigma_z^2}{\sigma_s^2}a^2\right)}\left(1+\frac{1}{k}\right) &\Leftrightarrow
\end{aligned}$$

$$2 - \frac{\gamma}{a} < \frac{\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2} a(2a-\gamma)}{\frac{\sigma_f^2}{\sigma_s^2} + \frac{\sigma_g^2 + b^2\sigma_z^2}{\sigma_s^2} a^2} \left(1 + \frac{1}{k}\right) \text{ and observe that}$$

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

which is indeed true.

To sum up, under the condition

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

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 $0 = \phi_1(a, e^*(a))$.

A.1.5 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 (1.7) 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 $\mathbb{E}_t\left[r_{t+1}|r_t, v_{c,t}, v_{s,t}\right]$.

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:

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

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

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

The projection theorem for multivariate Normal distributions implies:

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

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

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

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

After straightforward algebra, one finds that

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

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

Assembling altogether:

$$\mathbb{E}\left[x|y,|v|,u\right] \approx \left(\beta_{xy} + \rho\beta_{xv}|v|^2\right) y + \beta_{xu}u.$$

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

$$\mathbb{E}\left[x|y,|v|,|u|\right] \approx \left(\beta_{xy} + \rho_{yv}\beta_{xv}|v|^2 + \rho_{yu}\beta_{xu}|u|^2\right) y,$$

which is the analogue of (1.10). 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:

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

A.2 Data and sample

A.2.1 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.

A.2.2 Actively traded CDS contracts

DTCC publishes a list of 1000 most actively traded single-name CDS contracts quarterly since June 2009.[2] 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 on CDS transactions like the total number of clearing dealers and average daily notional amount. In this chapter, 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.

A.2.3 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%;

[2] See DTCC website.

- 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%.

A.3 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 A.1. Summary statistics of the unfiltered bond-day panel. This is a counterpart of Table 1.1 that shows how sample characteristics change in the full unfiltered bond-day panel (no restriction on the number of days since the previous trade).

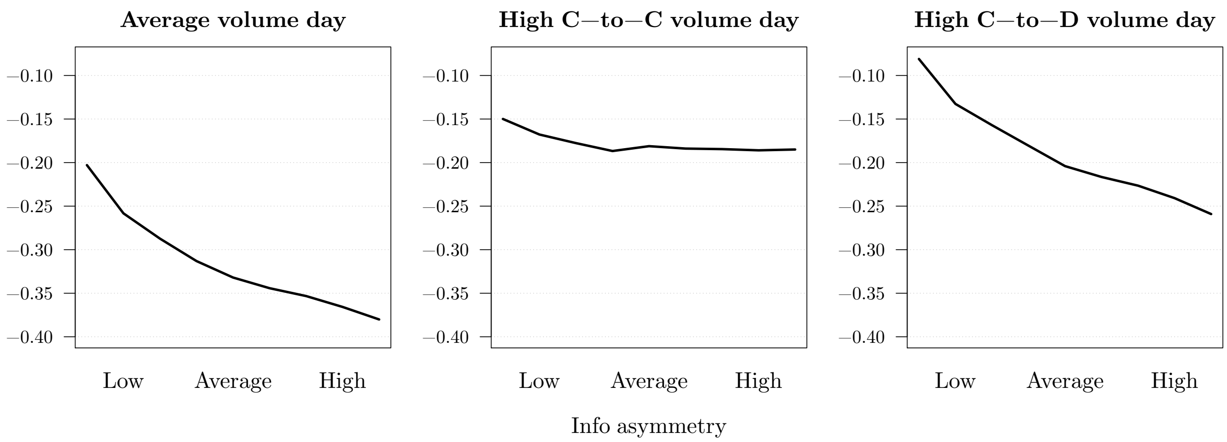


Figure A.1. Point estimates for high-volume day reversals. The calculations are based on models (8) from Tables 1.6-1.8. On the x-axis from left to right are the deciles of information asymmetry proxies. For instance, 'Low asymmetry' bond is the one that has the number of fund owners, active CDS dummy, 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	HY	IG	HY	IG	HY
	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\beta}_3$	
Intercept	-0.438^{***}	-0.515^{***}	0.131^{***}	0.134^{***}	0.068^{***}	0.090^{***}
	(0.010)	(0.028)	(0.013)	(0.029)	(0.010)	(0.026)
Average bid-ask	-0.076^{***}	-0.069^{***}	0.015^{**}	-0.010	-0.034^{***}	-0.042^{***}
	(0.006)	(0.011)	(0.006)	(0.011)	(0.005)	(0.010)
No. funds	0.002	0.012^{**}	-0.004	-0.006^{*}	0.001	0.002
	(0.002)	(0.005)	(0.002)	(0.004)	(0.002)	(0.004)
CDS dummy	0.002	-0.004	-0.003	0.002	0.002	-0.004
	(0.001)	(0.004)	(0.002)	(0.004)	(0.001)	(0.004)
Issue size	0.045^{***}	0.059^{***}	-0.008^{***}	-0.017^{*}	0.002	0.009
	(0.004)	(0.013)	(0.003)	(0.010)	(0.002)	(0.010)
No. dealers	0.012^{***}	0.039^{***}	-0.013^{***}	0.003	-0.005^{*}	0.005
	(0.003)	(0.010)	(0.003)	(0.006)	(0.003)	(0.007)
Issuer size	0.011^{***}	-0.002	-0.0004	0.0004	-0.004^{*}	-0.026^{***}
	(0.002)	(0.007)	(0.002)	(0.006)	(0.002)	(0.005)
–Equity volatility	0.004^{*}	0.006^{*}	-0.010^{***}	0.003	0.008^{***}	-0.002
	(0.002)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Risk controls	YES	YES	YES	YES	YES	YES
Vlm correlations	YES	YES	YES	YES	YES	YES
Observations	3,971	710	3,971	710	3,971	710
R^2	0.440	0.381	0.040	0.045	0.097	0.116
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table A.2. 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.
$\hat{\beta}_1$	-0.3823	-0.3944	17	9745	0	9160	9762
$\hat{\beta}_2$	0.0651	0.0558	6919	2843	1600	201	9762
$\hat{\beta}_3$	0.0531	0.0504	6767	2995	1934	357	9762

Table A.3. 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 1.4, 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	-0.480^{***}	-0.487^{***}	0.116^{***}	0.115^{***}	0.044^{***}	0.049^{***}
	(0.006)	(0.007)	(0.008)	(0.009)	(0.006)	(0.007)
Average bid-ask	-0.037^{***}	-0.042^{***}	0.012^{**}	0.013^{**}	-0.037^{***}	-0.032^{***}
	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)
C-to-C vlm corr.	0.008^{***}	0.008^{***}	0.013^{***}	0.012^{***}	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
C-to-D vlm corr.	-0.007^{***}	-0.008^{***}	-0.014^{***}	-0.014^{***}	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
No. funds	0.016^{***}	0.014^{***}	-0.002	-0.002	0.006^{***}	0.004^{**}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
CDS dummy	-0.001	-0.001	-0.003^{*}	-0.003^{*}	0.002	0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Issue size	0.019^{***}	0.020^{***}	-0.011^{***}	-0.010^{***}	-0.002	-0.0001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
No. dealers	0.013^{***}	0.013^{***}	-0.009^{***}	-0.011^{***}	-0.005^{**}	-0.002
	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Issuer size		-0.003^{*}		-0.002		-0.006^{***}
		(0.002)		(0.002)		(0.002)
–Equity volatility		-0.011^{***}		-0.009^{***}		0.004^{**}
		(0.002)		(0.002)		(0.002)
Risk controls	YES	YES	YES	YES	YES	YES
Observations	4,985	4,656	4,985	4,656	4,985	4,656
R^2	0.256	0.270	0.030	0.036	0.090	0.087
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table A.4. 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.

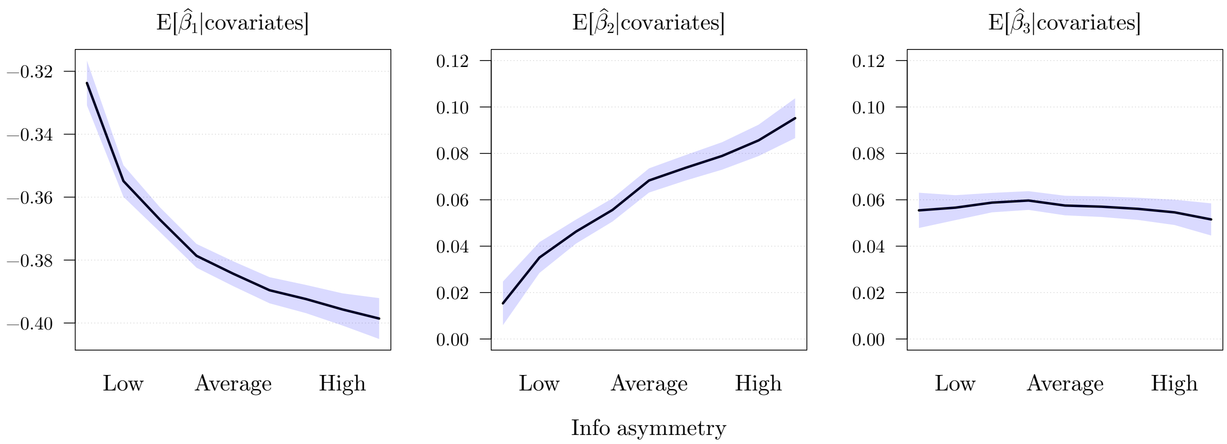


Figure A.2. Point estimates and confidence intervals for the expected values of volume-return coefficients (market return included in the first-step regression). This figure is a counterpart of Figure 1.3, but the volume-return coefficients are estimated controlling for market return in the first-step regression (see Table A.3).

	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 A.5. 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 1.4, 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	−0.417***	−0.440***	0.186***	0.187***	0.071***	0.078***
	(0.007)	(0.008)	(0.012)	(0.013)	(0.007)	(0.008)
Average bid-ask	−0.073***	−0.082***	0.021***	0.020**	−0.044***	−0.039***
	(0.005)	(0.005)	(0.007)	(0.008)	(0.004)	(0.005)
C-to-C vlm corr.	0.012***	0.011***	0.003	0.003	−0.002	−0.002
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
C-to-D vlm corr.	−0.009***	−0.009***	−0.012***	−0.009*	0.002	0.001
	(0.002)	(0.002)	(0.004)	(0.005)	(0.002)	(0.002)
No. funds	0.006***	0.006***	−0.008***	−0.008***	0.006***	0.003
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
CDS dummy	0.001	0.001	−0.002	−0.002	0.003*	0.001
	(0.001)	(0.001)	(0.003)	(0.003)	(0.001)	(0.001)
Issue size	0.041***	0.037***	−0.016***	−0.013***	−0.006**	−0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
No. dealers	0.018***	0.021***	−0.014***	−0.014***	−0.004	−0.001
	(0.003)	(0.003)	(0.004)	(0.004)	(0.002)	(0.003)
Issuer size		0.010***		−0.007**		−0.008***
		(0.002)		(0.003)		(0.002)
–Equity volatility		0.005**		−0.005		0.005**
		(0.002)		(0.003)		(0.002)
Risk controls	YES	YES	YES	YES	YES	YES
Observations	5,018	4,691	5,018	4,691	5,018	4,691
R^2	0.363	0.382	0.031	0.030	0.089	0.089
Note:	*p<0.1; **p<0.05; ***p<0.01					

Table A.6. 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.

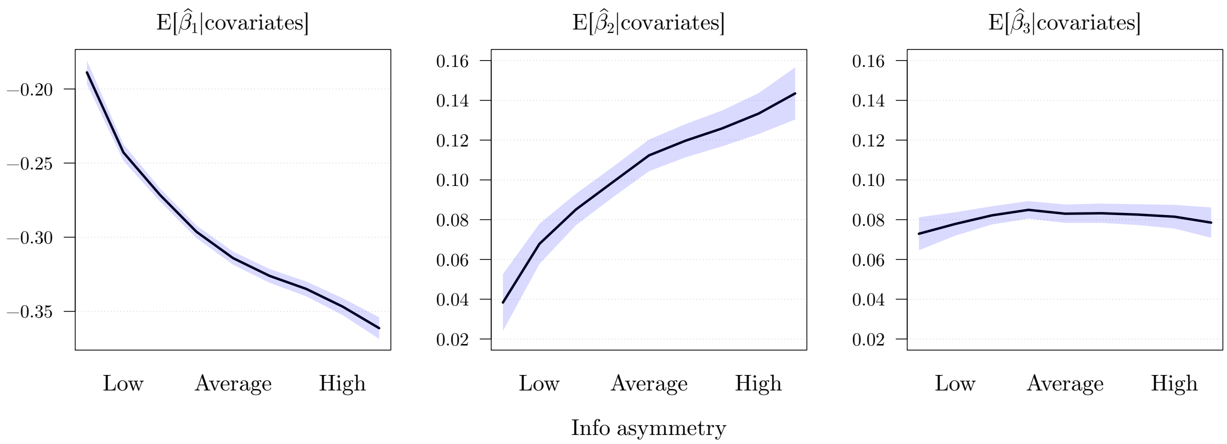


Figure A.3. Point estimates and confidence intervals for the expected values of volume-return coefficients (volumes included in the first-step regression). This figure is a counterpart of Figure 1.3, but the volume-return coefficients are estimated controlling for trading volumes in the first-step regression (see Table A.5).

Appendix B

Appendix of Chapter 2

B.1 Constructing the Dataset

B.1.1 Data Sources

This appendix describes step by step how the sample is constructed. Since I work in this chapter on the daily frequency, sample construction methodology differs in certain aspects from that for the monthly frequency.

Step 1. I start with the Enhanced TRACE intra-day bond market transactional data. The main difference of the Enhanced TRACE from regular TRACE is no cap on the reported transaction volume. This comes at a cost of a reporting lag. As of spring 2017 the Enhanced TRACE data are available through the WRDS only till the end of 2014, while plain TRACE data are available till the end of 2016. For the purpose of this study, it is not critical, though, to work with the most recent data; to have data on the exact transaction volume is more important. Full transaction volume allows me to compute bond liquidity measures. To ensure representativeness of the data I look at the so-called 'Phase 3' of the TRACE only (from October 2004 onwards). For detailed quantitative comparison of different phases of the TRACE see Asquith, Covert, and Pathak (2013b).

The Enhanced TRACE data needs to be cleaned of trade cancellations, reversals, corrections and agency transactions. The cleaning procedure I follow is described in Dick-Nielsen (2014). I also apply price filters to the data. All transactions with reported bond prices below 1 or above 500, as well as transactions with absolute returns above 20% (to a previous trade) are removed. Then I compute a daily average volume-weighted bond price and daily liquidity measures for each bond.[1] From this point onwards I work with the daily data.

Step 2. To obtain characteristics of the bonds I match securities from TRACE with Mergent FISD by CUSIP numbers. Once this is done, I reduce the sample to only **non-convertible senior unsecured corporate bonds** with less than 30 years to maturity. Callable bonds are not removed from the sample, but remaining outstanding amounts are tracked thanks to the file with historical outstanding amounts that is available in Mergent FISD.

Next, I determine for each bond for each day the exact remaining coupon payment/principal repayment schedule. This allows to compute daily prices of risk-free counterparts of the bonds by discounting remaining cash flows with Treasury zero-coupon rates for each particular day.[2] Then, observed bond prices and the prices of their risk-free counterparts are converted into yields to maturity. The difference in yields to maturity is the GZ spread (after Gilchrist and Zakrajšek, 2012).

I also obtain the history of credit rating revisions from Mergent FISD and add credit ratings to the sample. Throughout the chapter I use a numerical rating scale: 1 corresponds to 'AAA', 2 corresponds to 'AAA-', and so on, up to 22 that corresponds to 'D'.

Step 3. In this step, I add issuing firms' characteristics to the data. For this purpose, I match the issuers with the firms in CRSP and Compustat. By matching on either tickers,

[1] I re-did the study for simple daily average prices and daily last prices; it doesn't affect the results.

[2] These historical yield curves constructed as in Gürkaynak, Sack, and Wright (2007) are readily available via Quandl, `https://www.quandl.com/`. Here, the yield curve construction method is a modified Nelson-Siegel approach with additional parameters included to better fit the long end of the curve – the so-called 'Nelson-Siegel-Svensson' method.

or trade symbols, or 6-digit CUSIP numbers I am able to get the characteristics of issuing firms for more than 95% of the bonds (the rest are removed from the sample).

The ultimate goal of this step is to compute the Merton (1974) distance-to-default variable for each issuing firm for each day. For that I need firm equity value, volatility, and indebtedness for each day (see computational details in Appendix B.1.2). I obtain equity characteristics on the daily frequency from CRSP. Equity volatility is computed as the standard deviation of daily returns in the one-year rolling window. Firm indebtedness for each day is the latest available quarterly observation from Compustat carried forward. The default threshold needed to compute the distance-to-default is defined as all short-term debt and half of the long-term debt. Since the distance-to-default is a numerical solution to a system of equations, I remove bond(firm)-day observations for which this system doesn't have a solution with reasonable starting values.

Step 4. This is the step when I apply a number of final filters to the bond-day data. Below is the list of criteria according to which I remove observations from the sample. I remove:

– issuing firms from the financial and the real estate industry;

– bonds with less than one year to maturity;

– days with abnormally few trades;[3]

– observations in the 1st and the 99th percentiles of daily total returns;

– observations with the GZ spread below 5 b.p. or above 35%;[4]

– observations in the 99th percentile according to the Amihud illiquidity measure.

[3]These are the days with total number of trades per day at least 20% lower than the average daily number of trades over a 30-day rolling window. This criteria is reverse engineered – it allows to remove pre-holiday trading days. Cross-sectional distributions of bond prices and spreads on these days were found to be very different from the ones on regular days.

[4]Same filter as in Gilchrist and Zakrajšek (2012).

B.1.2 Daily Measure of Merton's Distance-to-default

In the Merton (1974) model firm's default probability at time t is determined by:

$$\mathbb{P}\left[V_A \leq D\right] = \Phi\left(-DD\right) = \Phi\left(d_1\right) = \Phi\left(-\frac{\log\left(\frac{V_A}{D}\right) + \left(r - \frac{\sigma_A^2}{2}\right)(T-t)}{\sigma_A\sqrt{T-t}}\right),$$

where V_A is the value of firm's assets, D is the default threshold, σ_A is the volatility of V_A, $T-t$ is the time to maturity, r is the discount rate, and $\Phi(\cdot)$ is the standard normal c.d.f.

To compute the DD variable one needs to know V_A and σ_A that are unobserved (unlike other parameters). There exist multiple methods to estimate these parameters, see Duan and Wang (2012) for a detailed overview. In this chapter, I use the 'volatility restriction' method that consists in solving for V_A and σ_A the following system of equations:

$$\begin{aligned}
0 &= V_A\Phi(d_1) - \exp\left\{-r(T-t)\right\} D\Phi(d_2) - V_E, \\
0 &= \frac{V_A}{V_E}\Phi(d_1)\sigma_A - \sigma_E,
\end{aligned}$$

where $d_2 = d_1 - \sigma_A\sqrt{T-t}$, and V_E and σ_E are correspondingly the value of the firm's equity and its volatility (these parameters are observed). I didn't use the transformed-data MLE approach to estimate the distance-to-default and opted for the volatility restriction method instead in order to speed up the computations. Solving numerically the system of equations above is orders of magnitude faster than running MLE estimations for each firm for each day. Experiments on a small sub-sample of the data didn't give considerably different results for the two methods.

In Section 2.2, I solve for V_A and σ_A for each firm for each day. V_E is the value of firm's equity from CRSP. σ_E is the standard deviation of daily equity returns from CRSP estimated over a backward-looking one-year long window. D is all short-term debt (less than one year to maturity) plus half of the long-term debt. Starting values for the solution algorithm are always $V_A[0] = V_E$ and $\sigma_A[0] = \sigma_E$. I disregard all firm-days when this approach doesn't lead to a reasonable solution.

B.2 Explanatory Variables

Here is the list of explanatory variables used in Sections 2.3–2.5.

- DD, the distance-to-default computed as presented in Appendix B.1.2. The values presented in Table 2.1 and further used in the analysis are scaled by 1000.
- DUR, the Macaulay duration.
- PAR, outstanding amount of a bond issue in mln USD. This variable contains the history of changes in the outstanding amount for each bond; a corresponding historical file is included in Mergent FISD (available through the WRDS server).
- CPN, a coupon rate of a bond, in % per annum.
- AGE, time elapsed since a bond was issued, in years.
- $CALL$, a call option dummy; equals to 1 if the bond issue is redeemable and to 0 otherwise.
- ADS, Aruoba et al. (2009) daily aggregate activity index for the US computed by the Philadelphia Fed and available (with historical vintages) at `https://goo.gl/mZJ5Sj`. This is a smoothed business cycle signal derived from 6 real activity series of different reporting frequency: weekly initial jobless claims, monthly payroll employment, industrial production, personal income less transfer payments, and manufacturing and trade sales, and quarterly real GDP. Since the index is obtained by running the Kalman smoother, its historical paths change a little bit as the new data become available.
- AMH, the Amihud liquidity measure, computed as presented in Section 2.2.
- LEV, an empirical proxy for the 'level' of the yield curve: 10-year zero-coupon rate y_{10Y}.

- SLP, an empirical proxy for the 'slope' of the yield curve: the difference between 10-year and 3-month zero-coupon rates $y_{10Y} - y_{3M}$.
- CRV, an empirical proxy for the 'curvature' of the yield curve: $2y_{2Y} - y_{10Y} - y_{3M}$.
- VOL, volatility of the long-rate: the standard deviation of the 10-year zero coupon rate computed over the 30-day rolling window.
- RFF, *Real Federal funds rate*, the difference between nominal effective Federal funds rate and realized (one month prior to a rate observation) 12-month CPI growth rate.
- TS, *Term spread*, same as the yield curve slope SLP.
- *GZ spread*, the corporate bond spread computed as in Gilchrist and Zakrajšek (2012).
- *Fitted GZ spread*, the portion of GZ spread explained by one the models of Table 2.2.
- EBP, excess bond premium, the difference between GZ spread and Fitted GZ spread.
- DRF, the default risk factor of Bai et al. (2019). This is the value-weighted average return difference between the highest-VaR quintile portfolio and the lowest-VaR quintile portfolio within each rating quintile portfolio. VaR is computed at the 5% level. For daily-rebalanced portfolios VaR is computed over the latest 100 days, for monthly-rebalanced portfolios over the latest 36 months.
- CRF, the credit risk factor of Bai et al. (2019). This is the value-weighted average return difference between the lowest-rating quintile portfolio and the highest-rating quinitile portfolio within each illiquidity quintile portfolio. Illiquidity portfolios are formed using the Amihud measure, unlike Bai et al. (2019), who use negative covariance between daily price changes as a low-frequency illiquidity proxy.
- LRF, the liquidity risk bond pricing factors of Bai et al. (2019). This is the value-weighted average return difference between the highest-illiquidity quintile portfolio and the lowest-illiquidity quintile portfolio within each rating quintile portfolio.
- SMB, HML, and UMD, are, correspondingly, small-minus-big, high-minus-low, and momentum stock pricing factors. Available from the Ken French's database via Quandl at: `https://www.quandl.com/data/KFRENCH-Ken-French`.

B.3 Time fixed effects in regressions for spreads

Here I consider alternative specifications for log spread fitting models with the time fixed effect TFE_t included:

$$\log\left(S_{it}^{GZ}[k]\right) = \beta \cdot DD_{it} + (\text{Proxies for recovery rate and liquidity}) + (\text{Call adjustment}) +$$
$$+ \eta \cdot AMH_{it}[k] + (\text{Industry and rating FE}) + TFE_t + \epsilon_{it}[k].$$

Compared to specifications in Section 2.3, this specification replaces ADS_t with TFE_t. Otherwise, the models are identical. As d'Avernas (2017) discusses in his Appendix E, such specification provides unbiased parameter estimates, unlike the benchmark Gilchrist and Zakrajšek (2012) model. My goal here is to extract the time fixed effect and investigate to what extent it is explained by aggregate business activity as measured by the ADS index.

Table B.1 presents the estimated models with time fixed effects. The first two columns correspond to a simple option adjustment with the same call dummy for all callable bonds (as Models 1–3 in Table 2.2), the last two columns also control for the interactions of a call dummy with the yield curve and bond-specific factors (correspond to Models 4–6 in Table 2.2). Note that the time fixed effect improves the overall fit of the models (compared to the specifications with the ADS index in Table 2.2). The models in Table B.1 capture more than 80% of the variation of log spreads. The coefficients on the Amihud measure in Tables 2.2 and B.1 are very close. However, the coefficients on the distance-to-default are considerably lower when the time fixed effect is included, in line with d'Avernas (2017) arguments.

Estimated TFEs from four alternative models are almost identical, the left chart on Figure B.1 shows. I will work with the TFE from Model 4 of Table B.1 since this model is the closest analogue of my preferred Model 6 of Table 2.2. To investigate the relationship between the TFE and the ADS I first run a standard OLS of the TFE on the ADS and a constant on the daily sample from Oct 2004 to Dec 2014. Such model has an R^2 of 0.63.

	Dependent variable: $\log(Spread_{it}[k])$			
	(1)	(2)	(3)	(4)
$-DD_{it}$	0.342***	0.341***	0.106	0.105
	(0.050)	(0.050)	(0.069)	(0.068)
$\log(DUR_{it}[k])$	0.321***	0.311***	0.295***	0.282***
	(0.010)	(0.010)	(0.020)	(0.019)
$\log(PAR_{it}[k])$	−0.087***	−0.082***	−0.075***	−0.070***
	(0.011)	(0.011)	(0.019)	(0.019)
$\log(CPN_i[k])$	0.470***	0.472***	0.694***	0.691***
	(0.019)	(0.019)	(0.048)	(0.047)
$\log(AGE_{it}[k])$	0.005	−0.0002	−0.032	−0.036*
	(0.006)	(0.006)	(0.021)	(0.021)
$CALL_i[k]$	0.025	0.029	0.719***	0.726***
	(0.020)	(0.020)	(0.263)	(0.259)
$AMH_{it}[k]$		0.043***		0.044***
		(0.002)		(0.002)
$-DD_{it} \cdot CALL_i[k]$			0.283***	0.284***
			(0.062)	(0.060)
$\log(DUR_{it}[k]) \cdot CALL_i[k]$			0.033*	0.037**
			(0.017)	(0.017)
$\log(PAR_{it}[k]) \cdot CALL_i[k]$			−0.013	−0.014
			(0.020)	(0.020)
$\log(CPN_i[k]) \cdot CALL_i[k]$			−0.248***	−0.245***
			(0.049)	(0.049)
$\log(AGE_{it}[k]) \cdot CALL_i[k]$			0.041*	0.039*
			(0.021)	(0.021)
$LEV_t \cdot CALL_i[k]$			0.020*	0.020*
			(0.012)	(0.012)
$SLP_t \cdot CALL_i[k]$			−0.026**	−0.026**
			(0.012)	(0.012)
$CRV_t \cdot CALL_i[k]$			0.025	0.024
			(0.016)	(0.016)
$VOL_t \cdot CALL_i[k]$			−0.002	−0.001
			(0.083)	(0.082)
Time FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Credit rating FE	YES	YES	YES	YES
Observations	2,756,326	2,756,326	2,756,326	2,756,326
Adjusted R^2	0.835	0.836	0.836	0.838
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table B.1. Explanatory models for the bond k of firm i – day t panel of credit spreads for the entire sample (Oct 4, 2004 – Dec 23, 2014) with the time fixed effect included. The dependent variable is the log of GZ spread. DD is the distance-to-default, DUR is duration, PAR is amount outstanding, CPN is the coupon rate, AGE is time elapsed from issuance, and $CALL$ is a callable bond dummy. AMH is the Amihud liquidity measure. LEV, SLP, and CRV are correspondingly level, slope, and curvature yield curve factors, and VOL is the realized volatility of the 10-year rate (30-day moving average). See Appendix B.2 for the details on explanatory variables. All models include also industry (the first two digits of the NAICS code) and credit rating (22-grade numeric scale) fixed effects. Standard errors are clustered in both firm i and time t dimensions.

The explained portion of the TFE is presented on the right chart of Figure B.1. To be sure that the result is not spurious I also do the Johansen cointegration test for the TFE and the ADS and estimate a corresponding error-correction model.[5]

[5] Both the TFE and the ADS are $I(1)$ over 2004–2014 period.

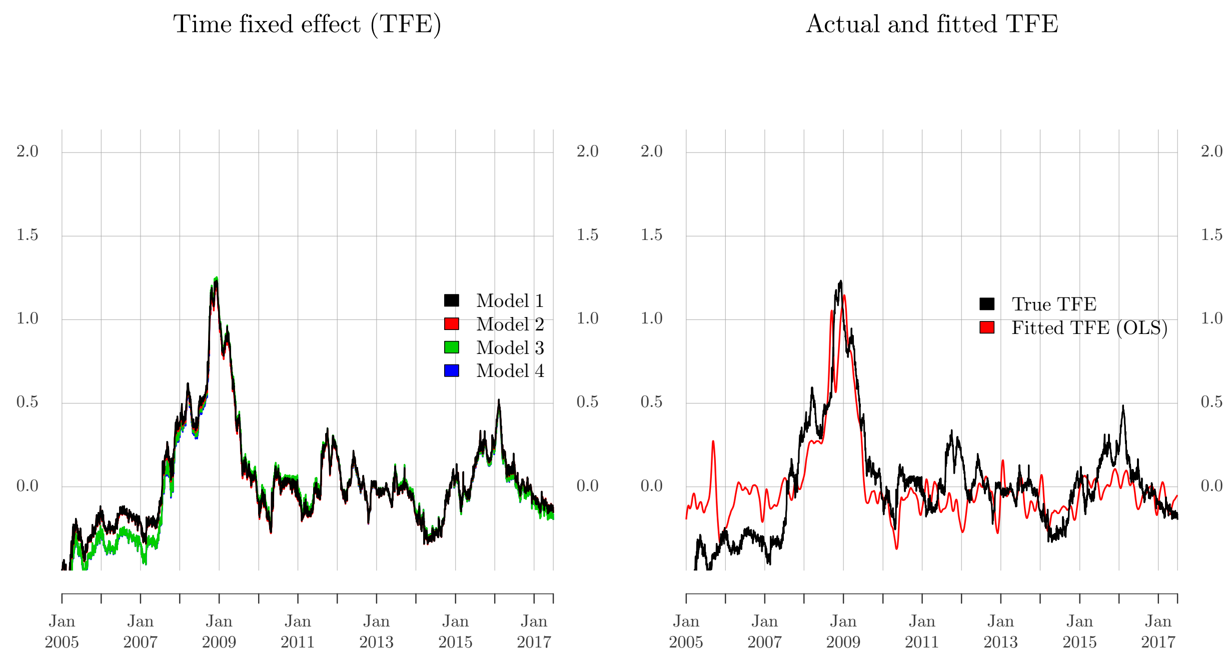


Figure B.1. Time fixed effect (TFE) extracted from the models of Table B.1. The left chart presents four alternative daily time-series of the TFE. The right chart plots the TFE from Model 4 vis-a-vis its fitted counterpart from the regression of the TFE on a constant and the ADS index. The sample is daily from Oct 4, 2003 to Dec 23, 2014.

Table B.2 demonstrates that the strong link between the TFE and the ADS is not spurious. The Johansen test (Table B.2a) rejects no-cointegration null at the 95% confidence level when more than two lags are included (the optimal number of lags is 15 according to the AIC). The estimated cointegration vector (Table B.2b) is statistically significant and economically reasonable. When the ADS drops from zero ('normal' times) to negative values (low activity states), the TFE jumps above its mean of 34 b.p.

In economic terms, the TFE absorbs time-varying portions of the remuneration for credit risk and of the credit risk premium. In this appendix, I demonstrated that this time-varying object is explained to a large extent by the aggregate business risk fluctuations as measured by the ADS index. This finding is in line with the results of Section 2.3 of the main text that emphasises aggregate business risk as the factor of credit spreads.

	Lag length			Critical values		
$\mathcal{H}_0$	2	3	15	90%	95%	99%
$r = 1$	2.54	2.99	4.60	7.52	9.24	12.97
$r = 0$	16.65	25.85	33.44	17.85	19.96	24.60

(a) Johansen cointegration test with trace-type test statistics. Lag length of 15 is optimal according to AIC. The null is in the leftmost column (r is the number of cointegration vectors). The null is rejected when the test statistics exceeds the critical value (the rightmost part of the table).

	TFE	ADS	Const
$\hat{\beta}^T$	1	0.43	0.14
	–	(8.45)	(3.29)
$\hat{\alpha}^T$	-0.01	-0.01	–
	(-3.19)	(-4.36)	–

(b) Cointegration vectors $\hat{\beta}$ and coefficients on the error-correction terms $\hat{\alpha}$ in the VECM with 15 lags. t-stats are in parenthesis.

Table B.2. Cointegration tests and vectors for the vector error-correction model (VECM) of daily TFE (Model 4 of Table B.1) and the ADS index. The sample is daily from Oct 4, 2003 to Dec 23, 2014.

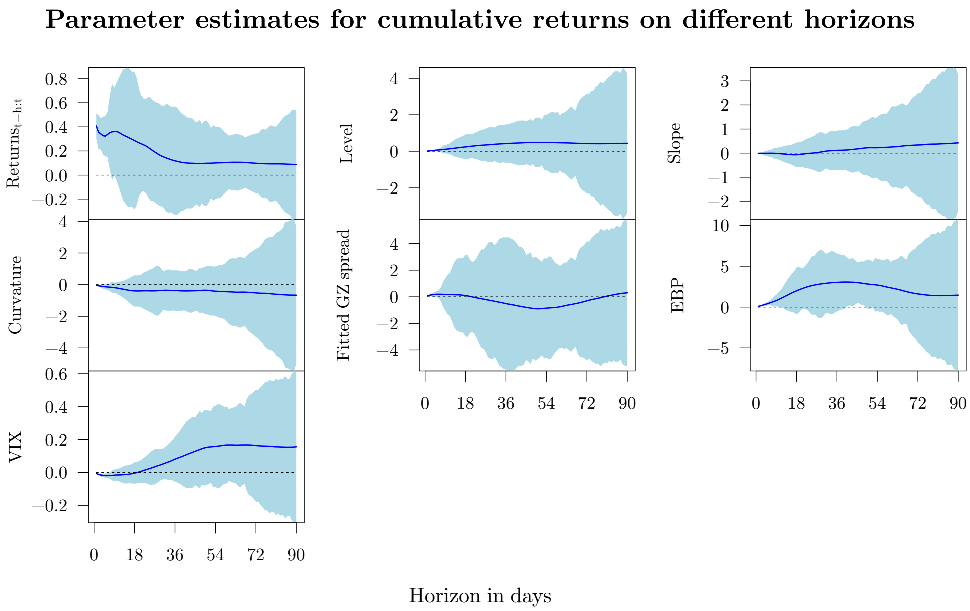
B.4 EBP and VIX as predictors of returns

Here I show that the predictive power of EBP for corporate bond market returns is immune to the inclusion of the VIX index in return forecasting regressions of Section 2.5.1. Estimation results are presented in Figure B.2. It is analogous to Figure 2.5, and the only difference is the VIX added to the right-hand side of forecasting models. On the top panel, Figure B.2a, the bond market is the TRACE portfolio of bonds, while on the bottom panel, Figure B.2b, the market is the Barclays Aggregate index.

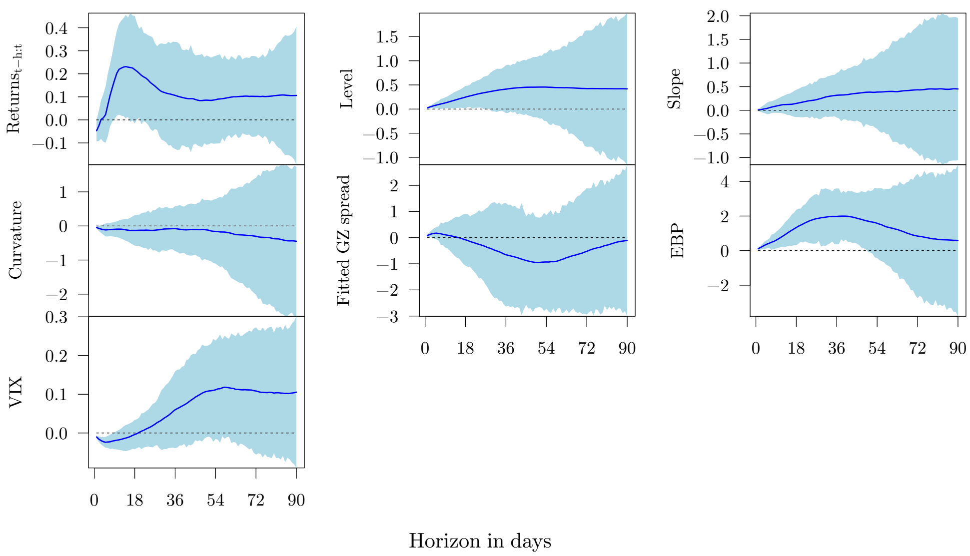
As Figure B.2 demonstrates, the VIX itself predicts market returns significantly only on horizons more than 50 days and only when the market is the TRACE portfolio. When the market is restricted to investment-grade bonds of the Barclays index only, the VIX is not significant on any horizon.

The coefficients on EBP remain significant for a wide range of forecasting horizons in the models with the VIX index added. When the market is the investment-grade index (Figure B.2b), EBP significantly predicts excess market returns on horizons up to several months, and the economic significance is only marginally lower than in regressions without VIX in Figure 2.5. This result applies to the TRACE portfolio as well, but here the addition of VIX compromises statistical significance on shorter horizons.

Daily VIX is a difference-stationary variable over the years 2004–2014, while the EBP is a level-stationary series. Replacing the levels of VIX in return-forecasting models by its first differences doesn't undermine the predictive power of EBP for market returns (not reported). Same applies to regressions with only the EBP and the VIX or its first differences on the right-hand side (also not reported). To sum up, even though the VIX might be a predictor of corporate bond market returns on some horizons, it doesn't stand behind the forecasting power of EBP for market returns.



(a) Dependent variable: returns on **TRACE portfolio** of bonds; fitted GZ spread and EBP of **Model 6** as explanatory variables.



(b) Dependent variable: returns on **Barclays Aggregate U.S. corporate bond index**; fitted GZ spread and EBP of **Model 6** as explanatory variables.

Figure B.2. Estimated forecasting regressions for **cumulative bond market excess returns**. Same explanatory variables as in Figure 2.5, plus the VIX index.

Appendix C

Appendix of Chapter 3

C.1 Sample selection

We apply the number of filters to the TRACE database *after* cleaning it as in Dick-Nielsen (2014) (we *do not* remove agency trades). Here are the criteria we use to select the sample:

- The trade was executed between Oct 4, 2004 and Dec 31, 2014;
- The bond is nominated in USD;
- Fixed coupon (including zero-coupon), non-asset backed, non-convertible, non-enhanced bond;
- 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;
- The quoting convention is 30/360;

- The trades are 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 don't fall into this range, they are removed from the sample);
- The trades are executed on NYSE business days;
- The bond was traded for at least two days in the sample period;
- The trade was executed on or after the dated date of the bond (the date when the interest starts to accrue).

C.2 SEC N-Q forms and holdings data

Mutual fund N-Q forms are available online through the SEC EDGAR system. We machine-read these forms and recover holdings from this scraped textual data. Mutual funds have a lot of discretion in how they fill their N-Q forms which makes the recovery of holdings difficult. We discuss the main steps we take below.

First and foremost, funds normally do not report bond CUSIP numbers in N-Q forms. Bond holdings in N-Q forms are identified by the issuer name, maturity, and coupon rate. Instead of trying to fill CUSIP numbers for all N-Q records we find N-Q records matching the CUSIPs we are interested in. We start with a list of CUSIPs from our sample (about 14 thousand as stated in Table 3.1), take their maturity, coupon rate, and issuer name; and match this dataset with N-Q records by maturity and coupon rate. Several possibilities arise. If there is no match, we remove such CUSIP from our 'NQ-matched subsample' (column 2 of Table 3.1).[1] If there is a match it may or may not be unique. Even if the match is unique (which is the dominant case observed for about 9 thousand bonds of interest), there is no guarantee that it is not some other bond, not from our plain-vanilla USD-denominated corporate bond sample, with the same coupon rate and maturity. To check that we compute a cosine text similarity measure between the true issuer name from the FISD database and

[1] An alternative way would be to assign the value of zero to mutual fund holdings of such bonds. Since funds rebalance infrequently, we do not want to overpopulate our sample with 0 changes in fund holdings.

an issuer name we recover from N-Q forms.[2] Table C.1 provides some examples. Table C.1a shows a record with a uniquely identified bond while Table C.1b shows a record with double matching: one bond is the true bond we are looking for, another bond is a mortgage-backed security with the same coupon and maturity. Regardless of whether the match is unique or not, we keep a record in our sample only if the 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 C.1. Examples of records with 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 an N-Q filing matched to the FISD data by maturity and coupon rate. 'Similarity' is a cosine similarity between 'issuer' and 'what' fields.

On the next step, we recover dollar holdings of the matched securities for every combination of bond–fund–reporting date. The raw data we have for every such observation is a string of dollar-like values, see an example in Table C.2. The most frequent case is when the par value and the market value are reported. We attempt to recover the par value, which is usually a number with a string of zeroes in the end. Sometimes funds also report the number of securities held (which is the par value divided by 1000 in almost all cases), together with the dollar par value or instead of it. Another complication comes from the fact that funds often scale dollar values in their reports by 1,000 or 1,000,000. In this case, the string that captures the table header contains a scaling unit, in numerical or textual form. We develop an algorithm that takes into account these and some other less frequent reporting patterns and recovers a dollar par value for every bond–fund–reporting date observation.

Given the nature of the data, we can not be sure that the algorithm recovers all holdings correctly. Because of that, we apply some additional checks and adjustments once we obtain

[2]We experimented with different similarity measures and did not observe much difference in results.

cusip_id	issuer	maturity	rate	report	CIK	dollars
22541LAL7	credit suisse first boston (usa) inc	2009-01-15	3.88	2005-01-31	0000933996	[365000, 362038]

Table C.2. An entry with dollar fields. Same entry as in Table C.1a. 'Dollar' field is a text string that contains all dollar values found in the row corresponding to the entry.

all candidate holding values. For instance, we track the holdings that are 'too high' relative to the outstanding amounts and scale them down assuming that we did not capture the scaling unit correctly at the previous step. Similarly, we scale down holdings that are unrealistically high relative to the average fund ownership in a given bond in a given month. We also truncate holdings at 1% and 99% in the entire bond–fund–reporting date sample; the tails are removed from the data.

In this chapter, we are interested in aggregated fund holdings per bond per month. Before aggregating holdings across funds we need to make an additional assumption about how funds rebalance their holdings. N-Q forms are submitted twice every fiscal year, which is fund-specific. So, funds report their holdings asynchronously. We test several ways of interpolating these data to the monthly frequency: 'last observation carried forward' (all rebalancing happens in the reporting month), linear interpolation (rebalancing in equal portions throughout six months), and exponential interpolation (more rebalancing in months right before the reporting months). In the chapter we present the results with the 'last observation carried forward' approach, they are qualitatively similar to the two other methods.

C.3 Additional tables and charts

	State 1	State 2	State 3	State 4	State 5
State 1	0.609	0.285	0.089	0.015	0.001
State 2	0.227	0.442	0.257	0.065	0.009
State 3	0.075	0.271	0.424	0.189	0.040
State 4	0.018	0.098	0.272	0.430	0.182
State 5	0.002	0.017	0.070	0.220	0.691

(a) (I)TB, pre-crisis

	State 1	State 2	State 3	State 4	State 5
State 1	0.674	0.246	0.070	0.010	0.001
State 2	0.249	0.454	0.239	0.052	0.006
State 3	0.093	0.314	0.414	0.153	0.027
State 4	0.029	0.149	0.334	0.352	0.136
State 5	0.002	0.018	0.064	0.145	0.770

(b) (I)TB, post-crisis

	State 1	State 2	State 3	State 4	State 5
State 1	0.900	0.071	0.022	0.006	0.001
State 2	0.199	0.395	0.263	0.113	0.030
State 3	0.032	0.139	0.391	0.306	0.132
State 4	0.004	0.031	0.158	0.426	0.381
State 5	0.000	0.002	0.015	0.083	0.900

(c) Non-(I)TB, pre-crisis

	State 1	State 2	State 3	State 4	State 5
State 1	0.920	0.059	0.016	0.004	0.001
State 2	0.228	0.403	0.240	0.104	0.025
State 3	0.035	0.134	0.396	0.314	0.121
State 4	0.006	0.036	0.192	0.430	0.336
State 5	0.000	0.002	0.021	0.097	0.879

(d) Non-(I)TB, post-crisis

Table C.3. Estimated monthly transition probabilities. State 1 is G1 of trading frequency ($Z \in [0, 20)$), state 2 is G2 ($Z \in [20, 40)$), etc. The underlying model is a five-state continuous time Markov chain with constant generator and instantaneous jumps to neighbouring states only.

	Mean	Median	S.D.	Min	5th	25th	75th	95th	Max	N.Obs.
					(I)TB					
Big trades	-0.03	0.00	2.19	-17.98	-2.78	-0.37	0.30	2.72	15.30	305763
Small trades	0.04	0.00	0.27	-5.72	-0.09	-0.02	0.04	0.26	9.49	305763
					Non-(I)TB					
Big trades	-0.02	0.00	1.81	-17.98	-1.88	0.00	0.00	1.80	15.30	651880
Small trades	0.02	0.00	0.32	-5.72	-0.06	0.00	0.01	0.16	9.49	651880

(a) Levels

	Mean	Median	S.D.	Min	5th	25th	75th	95th	Max	N.Obs.
					(I)TB					
Big trades	-0.01	0.00	3.18	-33.28	-4.22	-0.72	0.58	4.38	33.28	301877
Small trades	-0.00	0.00	0.27	-10.15	-0.18	-0.03	0.03	0.18	9.51	301877
					Non-(I)TB					
Big trades	-0.00	0.00	2.64	-33.28	-3.06	-0.11	0.02	3.05	33.28	641479
Small trades	-0.00	0.00	0.40	-15.20	-0.12	-0.01	0.01	0.11	15.20	641479

(b) Changes

Table C.4. Distribution of monthly levels and changes in net client buy volume conditional on trade size, in % of outstanding amounts. Volumes are winsorized at 0.001% and 0.999%.

Bond type	Mean	Median	S.D.	Min	5th	25th	75th	95th	Max	N.Obs.
					Mutual fund holdings					
(I)TB	12.29	9.20	10.78	0.00	0.66	4.41	16.78	40.29	42.08	280748
Non-(I)TB	10.78	7.28	11.00	0.00	0.34	2.82	14.29	42.08	42.08	455766
					Net purchases by mutual funds					
(I)TB	0.11	0.00	0.84	-19.98	-0.30	0.00	0.01	0.99	37.62	255010
Non-(I)TB	0.09	0.00	0.76	-26.12	-0.19	0.00	0.00	0.84	40.93	406569
					Net purchases by insurance companies					
(I)TB	-0.03	0.00	3.47	-70.35	-1.48	0.00	0.00	1.32	39.89	301874
Non-(I)TB	-0.17	0.00	4.45	-70.35	-0.82	0.00	0.00	0.69	39.89	636355

Table C.5. Distribution of mutual fund holdings, changes in holdings, and net purchases of insurance companies. Mutual fund (MF) holdings are analyzed for the subset of data that contains only bonds matched in SEC NQ filings. MF holdings are winsorized at 5% and 95%, changes in holdings are computed on the winsorized data. Insurance companies' (IC) net purchases are analyzed in the entire sample (all bond-month observations with no recorded purchases by insurance companies are filled with zeros). IC net purchases are winsorized at 0.1% and 99.9%.

	Dependent variable: ΔZ_{it}					
	(1)-(3) = (I)TB			(4)-(6) = Non-(I)TB		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.09***			−0.40***		
Δ(Amount outstanding)$_{it}$, %	−0.001	−0.001	−0.001	−0.05**	−0.05**	−0.04**
Δ(Credit rating)$_{it}$, notch	−0.28***	−0.20**	−0.20**	−0.28***	−0.25***	−0.26***
Age$_{it}$, % of maturity at issuance	0.01***	0.01***	0.01***	−0.003***	−0.002***	0.004***
Δ(No. bonds of same issuer)$_{it}$	−0.43***	−0.41***	−0.45***	−0.18***	−0.18***	−0.17***
Coupon month dummy$_{it}$	−2.57***	−2.78***	−2.84***	−1.48***	−1.62***	−2.15***
Month FE	NO	YES	YES	NO	YES	YES
Firm FE	NO	NO	YES	NO	NO	YES
Observations	283,532	283,532	283,532	422,353	422,353	422,353
Adjusted R^2	0.002	0.03	0.02	0.001	0.01	0.02
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard errors are clustered by the bond CUSIP.					

Table C.6. Panel models for monthly changes in trading frequency ΔZ_{it}. Models (1) and (4) are pooled OLS, the rest are fixed-effect panel models.

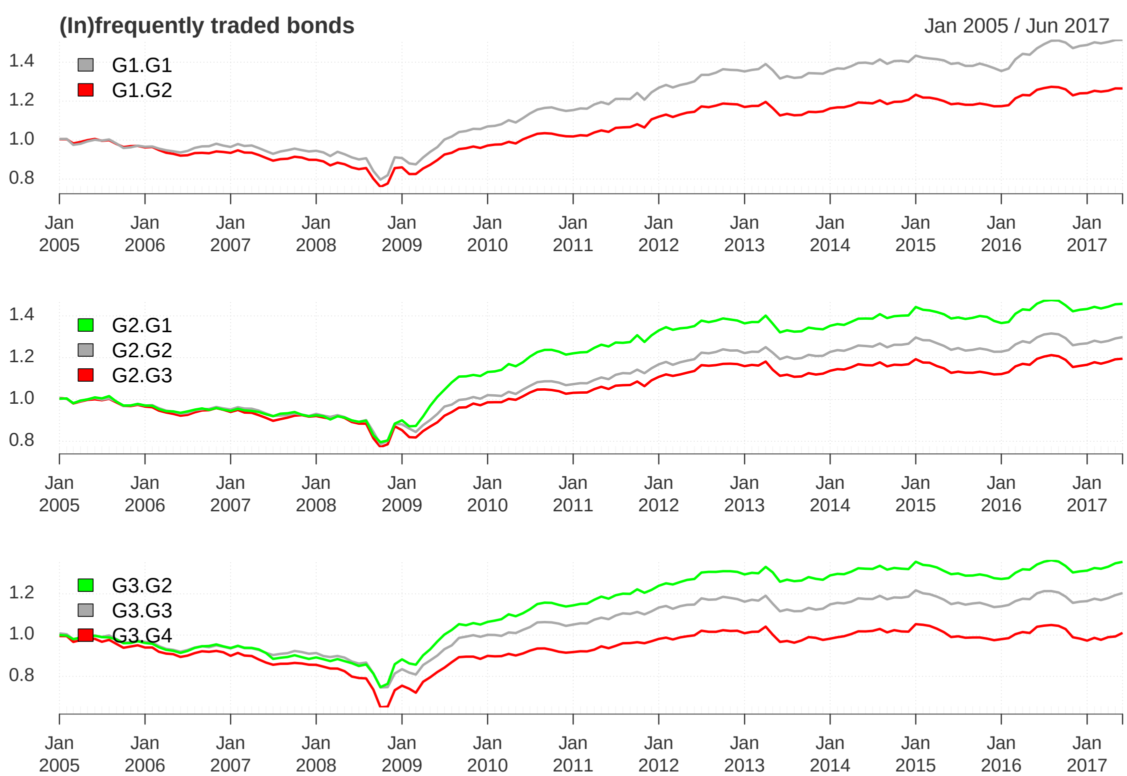


Figure C.1. Cumulative excess returns on (I)TB baskets based on the pairs of trading frequency. Excess returns are value-weighted returns in excess of the 3-month T-Bill rate. Baskets here are not investable since the trading frequency state in month T is not known apriori.

GR_{t-1}	GR_t	(I)TB R_t	(I)TB $Diff$	Non-(I)TB R_t	Non-(I)TB $Diff$
G1	G1	0.51**		0.25**	
G1	G2	0.23**	-0.28**	0.18**	-0.08*
G1	G3	0.24**	-0.26**	0.66**	0.41
G1	G4	-0.12	-0.63**	-0.64	-0.89*
G1	G5	0.41	-0.09	0.67	0.42
G2	G1	0.42**	0.22**	0.10*	-0.05
G2	G2	0.20**		0.15**	
G2	G3	0.15**	-0.06	0.24**	0.09
G2	G4	-0.03	-0.23**	0.10	-0.05
G2	G5	0.29	0.09	-0.02	-0.17
G3	G1	0.44**	0.32**	0.54*	0.38
G3	G2	0.27**	0.14**	0.27**	0.11*
G3	G3	0.13**		0.16**	
G3	G4	-0.13*	-0.26**	0.10**	-0.05
G3	G5	-0.28	-0.41	-0.12	-0.27**
G4	G1	1.06**	1.16**	0.49*	0.41*
G4	G2	0.47**	0.57**	0.35**	0.27**
G4	G3	0.21**	0.31**	0.21**	0.12**
G4	G4	-0.10		0.08**	
G4	G5	-0.38*	-0.28*	-0.10	-0.18**
G5	G1	1.28*	1.55**	1.48*	1.56**
G5	G2	1.83**	2.10**	0.79**	0.87**
G5	G3	0.87**	1.14**	0.32**	0.39**
G5	G4	-0.13	0.14	0.14**	0.22**
G5	G5	-0.27		-0.08	

Table C.7. Mean excess returns for the pairs of trading frequency groups in months $t-1$ and t. R_t is the mean return above the 3-month T-Bill rate in month t, $Diff$ is the difference in mean excess return relative to the case when a bond stays in the same trading frequency state in both months $t-1$ and t. **, and * correspond to 1%, and 5% significance.

	(I)TB	Non-(I)TB
G1-G1	0.196**	0.081
G1-G2	0.115	0.151
G1-G3	0.021	0.516**
G1-G4	-0.187	0.014
G1-G5	-0.202	1.242
G2-G1	0.167**	-0.020
G2-G2	0.141*	0.161
G2-G3	0.032	0.142
G2-G4	-0.003	0.193
G2-G5	0.288	-0.219
G3-G1	0.150*	0.500
G3-G2	0.197*	0.189
G3-G3	0.113	0.172*
G3-G4	-0.002	0.164
G3-G5	-0.414	0.368*
G4-G1	0.391	-0.019
G4-G2	0.272**	0.131
G4-G3	0.228*	0.215*
G4-G4	0.095	0.153
G4-G5	-0.058	0.035
G5-G1	0.378	0.695
G5-G2	2.075**	0.523*
G5-G3	0.454*	0.189
G5-G4	-0.003	0.139
G5-G5	-0.204	-0.178

Table C.8. Estimated alphas for bond portfolios formed by the pairs of trading frequencies in months $T-1$ and T. The underlying model is the Bai et al. (2019) corporate bond pricing model. 'Portfolios' here are not investable since the trading frequency state in month T is not known a-priori. Portfolio returns are computed by weighting individual bond excess returns by the market value of issues.