Managing Liquidity Risks in Bond Markets

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der KIT Fakultät für
Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Michael Reichenbacher

Tag der mündlichen Prüfung: 20. Mai 2021
Referentin: Prof. Dr. Marliese Uhrig-Homburg
Korreferent: Prof. Dr. Philipp Schuster
Karlsruhe 2021

Acknowledgements

I would like to express my deep gratitude to my supervisors, Prof. Dr. Marliese Uhrig-Homburg and Prof. Dr. Philipp Schuster, for their excellent guidance and continuous support throughout my doctoral studies. Our many interesting and inspiring discussions have been an immense asset for my research and this thesis. Also, I am very grateful to my co-author Prof. Dr. Melanie Schienle for her support and our insightful conversations. I also thank Prof. Dr. Steffen Rebennack for serving on my examination committee.

A special thanks to the second half of the junior research group, Jelena Eberbach, for the countless valuable and motivating conversations. Further, I am grateful to all my colleagues at the Institute of Finance, Banking, and Insurance. In particular, I thank Andreas Benz, Philipp Cölsch, Fabian Eska, Dr. Stefan Fiesel, Caroline Grauer, Dr. Martin Hain, Dr. Michael Hofmann, Annika Jung, Dr. Stefan Kanne, Viktoria Klaus, Dr. Marcel Müller, Anian Roppel, Dr. Meik Scholz-Daneshgari, Dr. Jan-Oliver Strych, Jun.-Prof. Dr. Julian Thimme, and Christian Wild for many stimulating conversations and for making the past years such an enjoyable time.

I gratefully acknowledge financial support by the Fritz Thyssen Foundation and the Deutsche Forschungsgemeinschaft (Grant No. UH 107/3-2 and SCHU 3049/2-2).

Lastly and most importantly, I want to express my heartfelt gratitude to my family for their love and endless support. Having you in my corner, failing is impossible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Trading frictions such as transaction costs play a fundamental role for investors in financial markets. These frictions have a particularly strong impact on bond markets due to bonds being one of the most traded assets. In 2020, market participants in the United States purchased and sold on average each day bonds with a nominal value of more than 950 billion USD, dwarfing the daily trading volume in US equities of about 480 billion USD.[1] However, while the digitization of exchanges over the last decades, e.g., with the transition from floor trading to electronic trading, led to diminishing trading costs in equity markets, investors still face a rather opaque and costly environment in most bond markets because of the markets' structure. Typically, bond markets are organized via a decentralized dealer network as over-the-counter markets. I.e., an investor trying to execute a transaction has to search for a dealer taking the counterpart and the price she pays or gets materializes as the result of the bilateral customer-dealer negotiation.[2] As a result, managing liquidity risks throughout each stage of an investment strategy is crucial for bond market investors and often determines the strategy's success. For example, during the initial set up, an investor needs to negotiate favorable quotes to minimize her initial costs but also to compensate for the transaction costs she expects to incur when liquidating the position (see, e.g.,

[1] For data on trading volume for 2020 on US bond markets see sifma (2021b) and on US equity markets see sifma (2021a).

[2] Electronic trading on bond markets has become more prevalent during the last years. While it has been mostly adopted on U.S. treasury markets with roughly 70% of volume being traded via e-trading platforms (see Greenwich Associates, 2018), it remains however rather an exception in other bond markets such as the U.S. corporate bond market where only between 19% to 26% of the volume in the first three quarters of 2018 was traded electronically (see Greenwich Associates, 2019).

Amihud and Mendelson, 1986). In a later stage, she needs to monitor the liquidity of her portfolio closely and adjust it accordingly to ensure that she can either react on negative market shifts or that the timely countermeasures prevent an unfavorable development in liquidity. Thus, for an effective management of these risks, investors are required to be able to accurately assess a bond's current liquidity as well as to anticipate possible changes in its future liquidity state.

The aim of this dissertation is to resolve issues that limit investors in their capability to manage liquidity risks by offering guidance to a more accurate liquidity measurement and a better understanding of liquidity effects on bond markets. In particular, throughout the thesis, we first introduce a liquidity measurement approach that incorporates the unique nature of over-the-counter bond trading. Second, we develop a sophisticated procedure to forecast individual bond liquidity as well as statistical tests to compare forecast models in such a setting. We exploit our new approaches in extensive empirical studies to reevaluate and gain further insights on liquidity effects in bond markets.

Measuring an asset's current liquidity is more involved for investors in over-the-counter bond markets than for assets traded on an exchange. In the absence of publicly available quotes and a limit order book, the literature has developed a battery of liquidity measures that exploit prices from past transactions and try to extract transaction costs from them (see Schestag, Schuster, and Uhrig-Homburg, 2016, for an overview of the most common measures). Working with past transactions however has the drawback that the past trading volumes usually do not coincide with the investor's actual trading demand during the initial set up of her strategy. For example, a retail investor exploiting data from large institutional trades faces the problem that she usually won't get as favorable quotes from her dealer as the past institutional investors. Due to the decentralized market structure, institutional investors are favored in the bilateral negotiations for several reasons. First, smaller investors are limited in the resources they can spend on the search for a counterparty for their trade while large investors can request multiple quotes and pick the most favorable offer. Second, once retail investors find a dealer they face fixed costs that are relatively more costly when trading a small volume (see Harris and Piwowar, 2006). And further, they lack the bargaining power of institutional investors because for example price discovery can not happen via quotes on public exchanges and thus bond investors are required to have a much better knowledge on the fair value of a bond (see, e.g., Green, Hollifield, and Schürhoff, 2007a,b). As a result, transaction costs diminish with increasing trading volume (see, e.g., Edwards, Harris, and Piwowar, 2007). Hence, if, as mentioned above, the trade sizes of the past transactions differ strongly from the

investor's desired trade size, the extracted trading costs are unlikely to be of value in the negotiations between the investor and her dealer.

The challenge imposed by trading costs depending on trade size is not only restricted to the initial set up of an investor's trading strategy but also transfers to monitoring the liquidity of her bond portfolio. Tracking changes over different months is usually done by estimating the average liquidity in the respective months. However, due to the cost-size dependence, this approach can be misleading in case that the average trade volume in the months differs strongly. In particular, changing average trade sizes are highly prevalent in two situations. The first case applies for the majority of bonds in the corporate bond market that are traded only a few times with randomly different trade volumes. If the random changes in trade size between months are sufficiently large, investors subsequently observe upward or downward jumps in average trading costs. And second, when assets are "fire sold", the market is flooded with large trades at high discounts (see, e.g., Feldhütter, 2012). In this case, investors observe an increase in average trade size but also a decrease in average trading cost due to the negative cost-size relationship. In this dissertation, we adapt the literature's standard liquidity measures to sever them from their purely mechanical link between trade volume and transaction costs. In addition, the adapted measures provide tailored guidance to investors on the average transaction costs to be paid for each desired trade demand.

Regarding the formation of expectations on a bond's liquidity in a future state, investors face the challenge that the literature so far does not provide guidance with an established forecasting model. Rather, researchers tend to use a bond's current liquidity as proxy in empirical applications when formally expected liquidity is required (see, e.g., Bao, Pan, and Wang, 2011; Dick-Nielsen, Feldhütter, and Lando, 2012, among others). This approach yet has various major shortcomings. First, a large part of bond investors do not pursue short-term strategies. As a result, it is highly questionable to assume that the liquidity of a bond is unaltered at the future selling date.[3] In fact, for example, bonds typically become more illiquid as they age (see, e.g., Jankowitsch, Nashikkar, and Subrahmanyam, 2011). Moreover, if investors assume that the liquidity of bonds will not change, they can only act on deteriorations in liquidity when they have occurred. Such behavior

[3]The inverse of the average yearly turnover in the U.S. corporate bond market during the time period from October 2004 to June 2017 is larger than one. Employing the inverse of the turnover as a proxy for an investor's expected holding period, i.e., that investors on average keep the bonds longer than a year in their portfolio. Further, note that bond investors usually have the option to hold until maturity to avoid liquidation costs. However, given the average time to maturity of roughly nine years in our sample, the expected holding period implies that a large part of investors indeed sell their positions before the bonds mature.

can lead to particularly strong losses in an investor's portfolio when liquidity dries up, e.g., during crisis periods. Acting late also has a negative impact when investors want to sell similar positions and the liquidation costs are internalized by the remaining investors. As a result, later trading investors are penalized with losses in the value of their position, which are further exacerbated when liquidity deteriorates (see, e.g., Goldstein, Jiang, and Ng, 2017). To overcome these shortcomings, we fill this gap and propose a dynamic forecast model for bond liquidity. Exploiting the model's predictions allows us to gain further insights into the strategic behavior of corporate bond fund investors.

From an econometric perspective, investors and researchers need appropriate statistical tests to compare competing forecast models they develop to form their expectations of a bond's future liquidity. Although there is a broad variety of studies developing test procedures to compare forecast models' performances in various settings (see, e.g., Diebold and Mariano, 1995; Clark and McCracken, 2001, for the general case or for nested models), these tests are based on a time series of forecasts for a single asset. However, in bond markets, we observe in general a large and diverse amount of issuers with several outstanding bonds. As a result of the additional heterogeneity in the cross-section, the standard tests are not suitable for comparing the accuracy of forecasts for individual bond liquidity. We analyze this issue and develop new statistical tests to provide researchers and practitioners with a tool kit to compare their forecast models in bond markets or in other financial markets with a vast cross-section.

1.2 Structure of the Dissertation

This dissertation is structured as follows:

In Chapter 2, which is based on the working paper Reichenbacher and Schuster (2020)[4], we focus on the challenges in OTC bond markets to accurately measure a bond's current liquidity when using past transaction data. To this end, we develop a new liquidity measurement approach for bond markets to overcome existing measures' weaknesses. As discussed above, the literature's measures suffer from the combination of two effects. First, transaction costs in OTC markets strongly depend on trade size (see, e.g., Edwards, Harris, and Piwowar, 2007). Second, many bonds trade only scarcely with strongly differing

[4]An earlier version of this paper was part of the habilitation thesis Schuster (2020). However, the current version differs significantly from this earlier version. While each section has been revised, the main difference lies in the new section 3 regarding the higher measurement precision of the new liquidity measures.

trading volumes. Therefore, changes in average transaction costs often indicate changing trade sizes rather than changes in the liquidity. We combine full-sample information for the size-cost relation with individual transaction data to eliminate such measurement problems. We show that our new measures are more precise in situations when less observations are available to calculate the liquidity measures, but also in asset fire sale situations when the average trade size increases systematically. Finally, exploiting the higher measurement precision, our size-adapted measures uncover the joint pricing of liquidity level and market liquidity risk in the cross-section of U.S. corporate bonds.

Chapters 3 and 4 focus on providing guidance for researchers and practitioners when predicting a bond's future liquidity. In Chapter 3, that builds on the working paper Reichenbacher, Schuster, and Uhrig-Homburg (2020), we introduce an approach to forecast individual bond liquidity and apply it to the U.S. corporate bond market. Our forecast approach is twofold. In each month, we start with possible predictors drawn from the literature on drivers of liquidity and identify the variables that have the strongest predictive power for the most recent past using three different statistical methods. We calibrate for each predictor set a linear model. Our final predictions are then the result of combining the predictions of these three dynamic models to obtain the most accurate estimate for future bond liquidity (see, e.g., Rapach, Strauss, and Zhou, 2010). We compare the new prediction methodology with the above discussed current literature's approach in empirical applications to use a bond's liquidity of today as the best estimate for its liquidity tomorrow. Our approach generates significantly lower forecasting errors and is much better able to capture the premium for expected liquidity in bond yields. We further provide evidence that investors in corporate bond funds actively anticipate liquidity deterioration in underperforming funds and sell their shares in advance to secure a first-mover advantage.

Chapter 4, which is based on the working paper Reichenbacher and Schienle (2021), addresses the challenges when testing for equal predictive accuracy of forecast models in the context of large panel data. The setting is characterized by forecast errors that exhibit strong heterogeneity across both dimensions the cross-section and the time series. We introduce new testing procedures that detect the heterogeneous structure in a clustering pre-step in a data-driven way. The standard errors in the Diebold and Mariano (1995) type test statistics are then adjusted accordingly based on this pre-step. We illustrate the new tests in the empirical setting of Chapter 3 and compare the predictive accuracy of different forecast models for individual bond liquidity. We show that without adequate control for two-dimensional heterogeneity in forecast errors, test results can be misleading or inconclusive.

Chapter 5 recapitulates the main results presented in this dissertation and gives an outlook on possible future research questions.

Chapter 2

Size-Adapted Bond Liquidity Measures and Their Asset Pricing Implications

2.1 Introduction

Quantifying transaction costs of bonds is important for investors, issuers, and regulators. Investors, for example, have to trade off the higher yield they get from illiquid bonds with the higher cost of trading. Regulators and central banks closely monitor the liquidity of a market and issuers have to pay higher yields to compensate for their bonds' illiquidity. Despite their importance, measuring a bond's transaction costs is difficult mainly due to the over-the-counter (OTC) nature of bond trading. If, for example, an investor wants to compare the costs of trading for two different bonds, she can request quotations from dealers. But due to large search costs, she is typically restricted to only a few observations. Moreover, requesting individual quotations is not feasible when monitoring the liquidity of a large bond portfolio. As an alternative, the investor can use a standard liquidity measure based on past transaction data. These measures typically assess a bond's costs of trading as average across all individual trades. Considering the strong dependence of transaction costs on trade size (see, e.g., Edwards, Harris, and Piwowar, 2007; Feldhütter, 2012), they depend heavily on the particular trade-size pattern observed for that bond. Thus, a fair comparison of the liquidity of different bonds or across time is in general not possible.

To illustrate this point, Panel A of Figure 2.1 depicts all buy and sell trade prices for a particular bond on one day. Using the average buy and sell price to calculate transaction costs, we get a relative bid-ask spread of roughly 2.5%. Now assume that, on the following day, the demand to trade large positions and the supply of liquidity stay identical but no investor has trading needs for volumes of less than \$500,000. On that day, the resulting bid-ask spread would be just 0.2% (see Panel B of Figure 2.1). So comparing those two days, an observer would mistakenly conclude that the liquidity of this bond improved strongly, although it did not change at all. Increases in the average transaction size might occur after a rating downgrade, when investors want to sell large positions, and bias observed liquidity changes around these events. Variations in the observed average trade sizes also occur idiosyncratically, especially for the majority of bonds that trade only a few times during a month. Some researchers try to address this problem and delete all transactions below a threshold, which is in most cases \$100,000. This approach has two major shortcomings. First, about two thirds of the transactions in the U.S. corporate bond market are retail-sized trades below \$100,000. As a result, large amounts of information are ignored, for which we show that they are valuable to accurately assess a bond's liquidity. Second, Figure 2.1 shows that the negative dependence of transaction costs on volume persists beyond \$100,000. Similarly, calculating transaction costs weighted by trade size strongly depends on the availability of large trades.

In this chapter, we introduce a two-stage approach to eliminate the size dependence of transaction costs in the bond market. In the first step, we estimate a market-wide function for the dependence of transaction costs on trade size. In the second step, we calculate a scaling factor for each bond and each month that scales the market-wide function so that it best explains the observed transaction costs for this bond. This factor can be interpreted as a measure of relative liquidity of the bond compared to a bond with average liquidity. Although the literature is aware of the relation between trade size and transaction costs in OTC markets, it is usually ignored when calculating common transaction cost measures. We are the first to incorporate this relation directly into liquidity measures, making them immune to changes in the individually observed trade-size pattern. Importantly, our new approach offers investors an easy way to calculate transaction costs for arbitrary position sizes as the product of the individual scaling factor and the market-wide transaction cost function evaluated at the trade size in question.[5]

[5]In the stock market, price impact measures like λ (see, e.g., Kyle, 1985; Hasbrouck, 2009) are typically employed to calculate size-dependent trading costs (see, e.g., Goyenko, Holden, and Trzcinka, 2009). Schestag, Schuster, and Uhrig-Homburg (2016) find that these measures do not work well in the bond market.

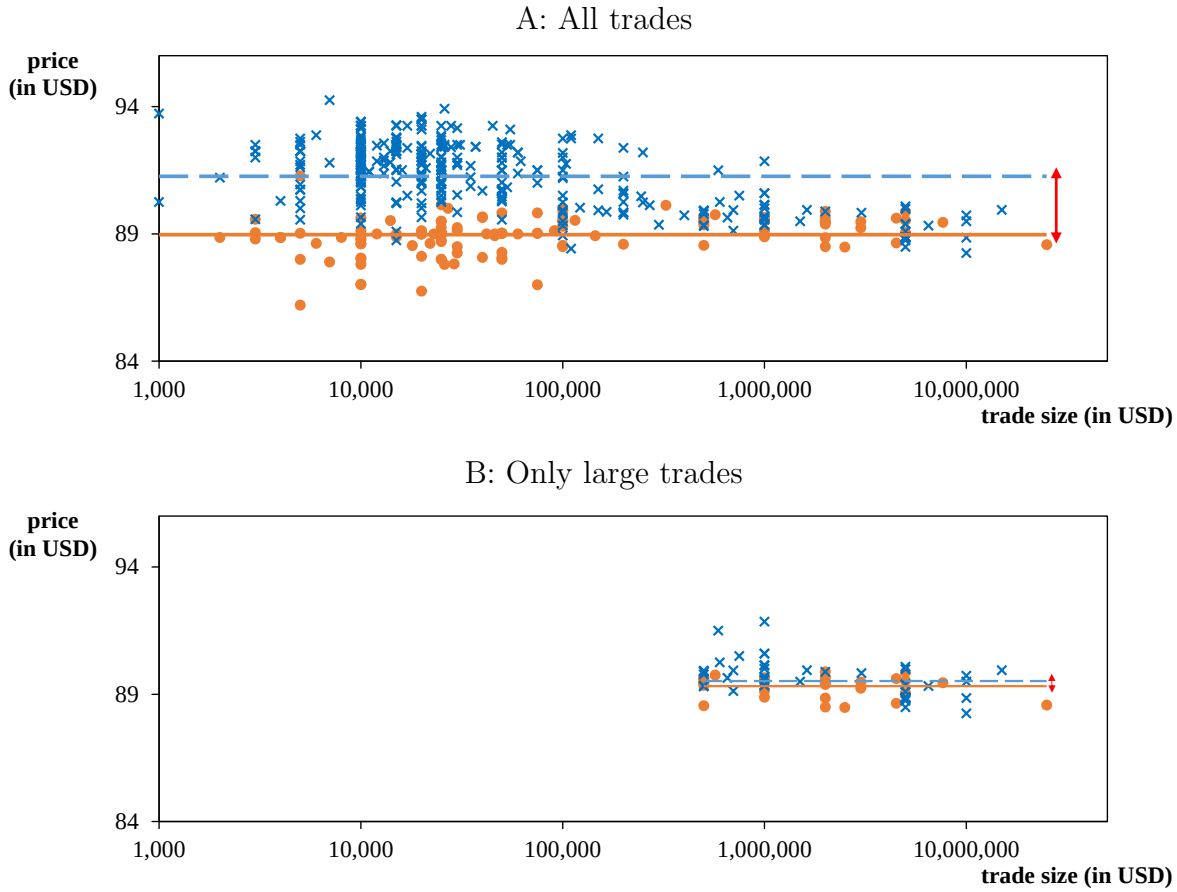


Figure 2.1: **Daily trades in a heavily traded bond**
This figure depicts an exemplary trading day in a U.S. corporate bond. The trading day is selected as the one with the largest number of observations for both buy and sell trades. The bond is issued by General Motors, has a fixed coupon of 8.375%, and matures in July 2033. The trading day is March 17, 2005. Blue crosses represent trades in which the customer buys (ask) and orange dots depict trades in which the customer sells (bid). The blue dashed line indicates the average buy price, the orange solid line the average sell price, and the red arrow the bid-ask spread. Panel A includes all available trades, while Panel B shows only trades with a trading volume larger than $500,000.

The combination of the full-sample information for the market-wide cost function with the individual transaction data to extract a bond's relative liquidity exploits all available information in a natural way. For that reason, our size-adapted measurement approach can be calculated with the same data requirements as standard measures when there is only one observation for transaction costs available. The new methodology is straightforward to implement and can be applied to adapt a large number of commonly used transaction cost measures. As different high-frequency liquidity measures are closely connected (see Schestag, Schuster, and Uhrig-Homburg, 2016), we exemplarily implement our approach for the Schultz (2001) and the average bid-ask spread measure (see, e.g., Hong and Warga, 2000), which are two established transaction cost measures. We use U.S. corporate bond transaction data from Enhanced TRACE for the period from October 1, 2004 to December 31, 2014.

We demonstrate that our size-adapted liquidity measures are associated with a higher measurement precision, which stems from two sources. First, the adapted measures are more accurate when there are less observations to calculate the liquidity measure. This situation occurs for the vast majority of corporate bonds that are only traded scarcely. In a bootstrap-like exercise, we analyze the error in the liquidity measures for highly traded bonds when a part of their trades is discarded. We find that our two size-adapted liquidity measures are on average associated with about 10% to 30% lower measurement errors. Hypothesizing that more precise liquidity measures will better explain the liquidity-related component of bond yield spreads, we run a second test. In a panel setting inspired by Friewald, Jankowitsch, and Subrahmanyam (2012), we regress yield spread changes on the changes of individual bond transaction costs and a set of control variables. The part of the R^2_{adj} that can be attributed to transaction costs almost doubles when using our size-adapted measures compared to their standard counterparts. In a robustness analysis, we show that deleting small trades as well as a simple trade-size weighting perform worse than unadjusted standard measures. This result shows that the transaction costs of small trades contain valuable information.

Second, our size-adapted liquidity measures are more precise in fire sale situations when the average trade size increases systematically. Using rating downgrades to junk status as a natural experiment, we show that the new measures are able to better capture the liquidity deterioration associated with the resulting fire sale. Standard transaction cost measures exhibit a systematic bias in such stress events. As a result, they do not detect a systematic change in liquidity. This might be the reason why existing studies on such events often focus on price impact measures (see Bao, O'Hara, and Zhou, 2018) or

use non-standard liquidity measures (see Dick-Nielsen and Rossi, 2019, who use returns of liquidity providing dealers).

In the second part of the chapter, we show that ignoring the dependence of common liquidity measures on the trade-size pattern can have a strong impact on the results when studying the asset pricing implications of liquidity. Accordingly, our size-adapted liquidity measures help to reconcile conflicting findings of previous studies regarding the question whether the level of liquidity or the risk of changing market-wide liquidity is priced. Building on the predictions of Amihud and Mendelson's (1986) model, there is a large literature confirming the influence of a bond's liquidity on its expected return (see, e.g., Amihud and Mendelson, 1991; Chen, Lesmond, and Wei, 2007; Bao, Pan, and Wang, 2011). In contrast to that, the results for bond market liquidity risk are conflicting. Whereas studies like Lin, Wang, and Wu (2011), Dick-Nielsen, Feldhütter, and Lando (2012), and Bai, Bali, and Wen (2019) confirm that bonds with a stronger return sensitivity to market-wide liquidity shocks earn higher expected returns, Bongaerts, de Jong, and Driessen (2017) do not find a significant bond market liquidity risk premium. These authors develop an integrated asset-pricing model to simultaneously analyze the effects of liquidity level and market liquidity risk on ex ante expected bond returns. We find that due to the strong correlation of transaction costs and liquidity beta, it is very difficult to disentangle the effects of liquidity level and risk. Indeed, when using standard measures of liquidity, the observed pricing pattern is strongly affected by the way test assets are constructed or betas are estimated, explaining the conflicting results. In contrast, when using our more precise size-adapted liquidity measures, we consistently find that both liquidity level and bond market liquidity risk are priced in the cross-section of expected bond returns. Economically, the part of yield spreads related to liquidity level and risk is large and amounts to 0.8% to 0.9% p.a. for each of the two dimensions. More broadly, our work also contributes to the literature on the risk versus characteristics debate (see, e.g., Davis, Fama, and French, 2000 or Daniel, Titman, and Wei, 2001 for the stock market and Gebhardt, Hvidkjaer, and Swaminathan, 2005 for the corporate bond market). Specifically, we show that an imprecise measurement of the characteristic can cause misleading conclusions on the question whether the characteristic or the associated risk are priced.

The plan for the rest of the chapter is as follows. In Section 2.2, we develop our size-adapted measurement approach. In Section 2.3, we show in several situations that our newly developed size-adapted measures are more precise than their standard (unadapted) counterpart. In Section 2.4, we perform asset pricing tests to answer the question whether the level of liquidity or market liquidity risk are priced. Finally, Section 2.6 concludes.

2.2 Size-Adapted Liquidity Measures

In this section, we describe our size-invariant liquidity measurement approach. We use a two-stage procedure to estimate the (relative) liquidity of a bond in each month. We implement this approach for two established transaction cost measures in Sections 2.2.2 and 2.2.3, using bond data from Enhanced TRACE from October 1, 2004 to December 31, 2014 (see Appendix A.1 for details on the data).

2.2.1 Basic Measurement Approach

Measuring the liquidity of a bond using common high-frequency liquidity measures is usually based on average transaction costs in this bond. As illustrated in Figure 2.1, such averaging across different transaction sizes ignores the size-dependence of transaction costs and a fair comparison between the liquidity of different bonds or even between a bond's liquidity for different periods is not possible.[6] Researchers often address this problem by excluding small trades entirely from the analysis. We show in Appendix A.2 that using this approach, a daily or monthly liquidity measure often cannot be calculated for a particular bond at all. Moreover, we show that small trades contain valuable information on a bond's liquidity and the combination of institutional and retail-sized trades offers the most comprehensive information set.

We develop a two-step liquidity measurement approach that combines information from all trade sizes in a natural way. In the first step, we estimate a function $c(vol)$ capturing the dependence of transaction costs on the traded (notional) volume vol. This transaction cost function serves as a market-wide benchmark for all bonds, representing an average state of liquidity. To ensure the highest possible generality for the cost function, we employ a nonparametric estimation based on all observations of all available bonds in the sample.[7] We verify later that the general form of the transaction cost function does not change much over time and confirm that small changes have no significant impact. In the second step, the liquidity of a bond i in month t (or day t) is measured relative to this market-wide

[6]The negative relation is unchanged after the introduction of the Volcker rule (see, e.g., Adrian, Fleming, Shachar, and Vogt, 2017; Bessembinder, Jacobsen, Maxwell, and Venkataraman, 2018) and it also holds for trades executed via electronic trading platforms (see Hendershott and Madhavan, 2015).

[7]It is also possible to estimate the function using parametric functional forms, which are in general easier to compute (see Edwards, Harris, and Piwowar, 2007). As a robustness check in Section 2.5.2, we repeat our asset pricing analyses from Section 2.4 employing liquidity measures based on a parametric functional form.

and time invariant function. We use the (estimated) transaction costs $tc_{k,i,t}$ of all observed transactions k of bond i in month t. We then express the relative liquidity of bond i in month t by an individual scaling factor $sf_{i,t}$, estimated in the following regression

$$tc_{k,i,t} = sf_{i,t} \cdot c(vol_{k,i,t}) + \epsilon_{k,i,t}, \tag{2.1}$$

where $vol_{k,i,t}$ denotes the transaction volume of the k-th transaction and $\epsilon_{k,i,t}$ an error term. A liquid bond is characterized by a scaling factor below 1, which indicates that its transaction costs for all trade sizes are smaller than those of an average bond. Similarly, an illiquid bond has a scaling factor larger than 1. Because all scaling factors are based on the same market-wide transaction cost function, changing trade-size patterns for a particular bond do not affect the factor estimate. Thus, our measure allows for a direct comparison of the liquidity of different bonds or different observation months. Most importantly, the full information regarding a bond's liquidity across trade sizes can be expressed as one number that is easy to interpret.[8] Thus, we can use the scaling factor $sf_{i,t}$ as a measure for the liquidity of bond i in month t.[9]

The size-invariant measurement approach described by Equation (2.1) can be applied to a broad variety of common liquidity measures. We choose to adapt the transaction cost measure of Schultz (2001) and the average bid-ask spread measure (see, e.g., Hong and Warga, 2000) as both exploit information on a transaction's trade side.

2.2.2 Size-Adapted Schultz (2001) Measure

The first measure that we adapt to the new approach is the (relative) Schultz (2001) liquidity measure that is based on the model

$$\Delta_{k,i,t} = \alpha_{i,t} + c_{i,t} \cdot D_{k,i,t} + \epsilon_{k,i,t}. \tag{2.2}$$

Here, $c_{i,t}$ approximates the average (relative) half-spread of bond i in month t. $\Delta_{k,i,t}$ denotes the (relative) price deviation of the trade price to a consensus price for trade k. $D_{k,i,t}$ is a trade side indicator that equals 1 for customer buys, -1 for customer sells, and

[8] We have experimented with more sophisticated functional forms that allow, e.g., transaction costs for large trade sizes to vary (relatively) stronger than transaction costs for small trades (see, e.g., Anderson and Stulz, 2017). The higher complexity only increases the precision very moderately.

[9] Note that the interpretation as a liquidity measure would not be possible if the transaction cost function were estimated for each bond and/or each month separately (see, e.g., Edwards, Harris, and Piwowar, 2007).

0 if the trade is between two dealers. We obtain daily consensus prices from Bloomberg.

Following the intuition in (2.1), we replace in (2.2) the average half-spread with the market-wide cost function $c(vol_{k,i,t})$ of trading the transaction volume $vol_{k,i,t}$ multiplied by the bond-individual scaling factor $sf_{i,t}^{Schultz}$, leading to

$$\Delta_{k,i,t} = \alpha_{i,t} + sf_{i,t}^{Schultz} \cdot c(vol_{k,i,t}) \cdot D_{k,i,t} + \epsilon_{k,i,t}. \tag{2.3}$$

We can now interpret the scaling factor $sf_{i,t}^{Schultz}$ as a measure for the relative liquidity of bond i in month t.

To estimate Equation (2.3), it would be best to simultaneously solve for the transaction cost function $c(\cdot)$ and the scaling factors. However, as such a simultaneous optimization is numerically problematic, we implement an iterative two-stage weighted regression. In the first step, we estimate a nonparametric function $c(\cdot)$ from all observations in the sample using the scaling factor estimates from the previous iteration.[10] In the second step, we use the transaction cost function from the first step and estimate (2.3) for each month t and each bond i separately to obtain individual scaling factors $sf_{i,t}^{Schultz}$. We exclude negative bid-ask spreads with the constraint $sf_{i,t}^{Schultz} \geq 0$. We weight observations in both steps to ensure that different volume segments contribute equally to the market-wide transaction cost function and the scaling factors. In this spirit, we define eleven segments centered symmetrically around the individual trade sizes \$10,000, \$25,000, \$50,000, \$100,000, \$200,000, \$500,000, \$1 million, \$2 million, \$5 million, \$10 million, and \$20 million.[11] To put equal weight on each segment, we weight each trade with the inverse of the full sample number of observations in the respective volume segment. To ensure that the cost function represents a state of average liquidity, we rescale it after each iteration such that the average of all scaling factors equals 1. We iterate over steps one and two until convergence.[12] Note that the data requirements to estimate an individual scaling factor $sf_{i,t}^{Schultz}$ in (2.3), given a market-wide function $c(vol_{k,i,t})$, are identical to the standard Schultz (2001) measure in (2.2), i.e., the regression is identified whenever we have at least two trades with different trade sides $D_{k,i,t}$ and the corresponding consensus prices.

The resulting transaction cost function is depicted in Figure 2.2. Consistent with previous studies, we find a negative relation between transaction costs and trade size.

[10] We employ locally weighted scatterplot smoothing (LOESS) as the nonparametric regression method. In the first iteration, we set all scaling factors to 1.

[11] Our results do not depend on the exact specification of these segments.

[12] The iteration terminates when the average absolute difference between the factors of the current and the previous iteration is below 10^{-6}.

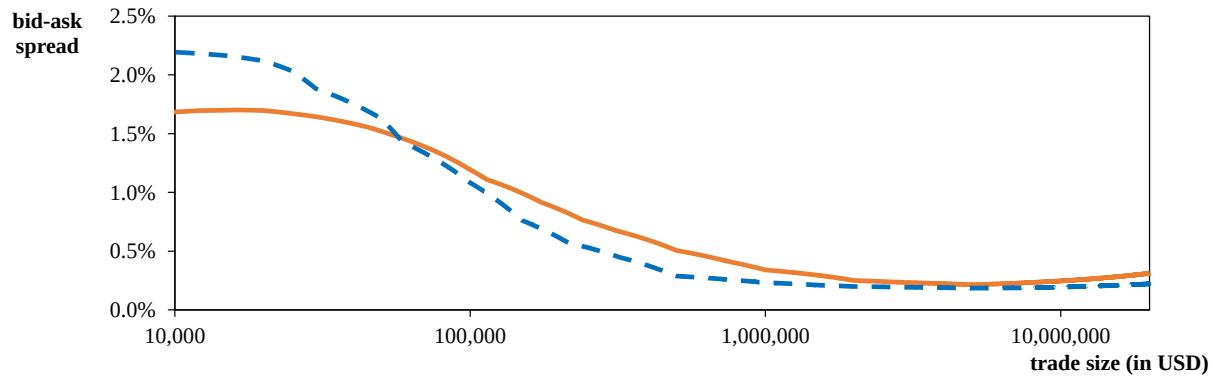


Figure 2.2: **Market-wide transaction cost functions of U.S. corporate bonds**
The figure depicts transaction costs dependent on trade size for the size-adapted Schultz (2001) measure (orange solid line) and the size-adapted average bid-ask spread measure (blue dashed line). The transaction cost functions are based on a nonparametric regression. They are estimated with the two-step measurement approach described in Sections 2.2.2 and 2.2.3.

For example, a retail investor who wants to trade a position of $10,000 pays on average round-trip costs between 1.5% and 2%. In contrast, an institutional position of $5 million trades at bid-ask spreads of only about 20 basis points. The literature mainly attributes this difference to the stronger negotiation power of institutional investors that is due to their lower search costs (see Feldhütter, 2012; Green, Hollifield, and Schürhoff, 2007b) or their more precise knowledge of a bond's fair value (see Green, Hollifield, and Schürhoff, 2007a). Fixed costs per trade might be playing a role as well (see Harris and Piwowar, 2006). For very large volumes though, we observe a slight increase in transaction costs. This increase likely comes from their higher inventory risk (Stoll, 1978) or difficulties in finding counterparties for these extremely large positions.

2.2.3 Size-Adapted Average Bid-Ask Spread Measure

The second measure that we adapt to our size-invariant measurement approach is the relative difference of average bid and ask prices

$$AvgBidAsk_{i,d} = \frac{\overline{P_{i,d}^{buy}} - \overline{P_{i,d}^{sell}}}{0.5 \cdot \left(\overline{P_{i,d}^{buy}} + \overline{P_{i,d}^{sell}}\right)}, \tag{2.4}$$

where $\overline{P_{i,d}^{buy}} = \frac{1}{n_{i,d}^{buy}} \sum_{k=1}^{n_{i,d}^{buy}} P_{k,i,d}^{buy}$ is the average customer buy price, $\overline{P_{i,d}^{sell}} = \frac{1}{n_{i,d}^{sell}} \sum_{k=1}^{n_{i,d}^{sell}} P_{k,i,d}^{sell}$

gives the average sell price, and $n_{i,d}^{buy/sell}$ is the number of buy/sell trade prices $P_{k,i,d}^{buy/sell}$ for bond i on day d, respectively. This measure was introduced by Hong and Warga (2000) and Chakravarty and Sarkar (2003). We can calculate it for each day with at least one buy and one sell trade. For a monthly measure, we calculate the mean of all daily bid-ask spreads.

Starting from our new approach (2.1), the adaption of this measure is slightly more involved than for the Schultz (2001) measure. Since the difference of average bid and ask prices is only available on a per day and not on a per trade basis, we have to evaluate the market-wide cost function for each trade that goes into the calculation of the daily average separately and then compute the average:

$$AvgBidAsk_{i,d} = sf_{i,t}^{AvgBidAsk} \cdot \frac{1}{2} \left[\frac{1}{n_{i,d}^{buy}} \sum_{k=1}^{n_{i,d}^{buy}} c(vol_{k,i,d}^{buy}) + \frac{1}{n_{i,d}^{sell}} \sum_{k=1}^{n_{i,d}^{sell}} c(vol_{k,i,d}^{sell}) \right] + \epsilon_{i,d}, \quad (2.5)$$

where we plug in daily average bid-ask spreads $AvgBidAsk_{i,d}$ from (2.4) as the left hand side of (2.5). We again scale the average of the market-wide transaction cost function with a bond-individual scaling factor $sf_{i,t}^{AvgBidAsk}$. Because, in general, the number of sell and buy trades on a given day is not identical, the underlying prices enter with differing weights into the calculation of $AvgBidAsk_{i,d}$. To match this imbalance, we calculate the average of the market-wide cost function $c(vol_{k,i,d}^{buy/sell})$ separately for both sides before computing the overall average. Again $n_{i,d}^{buy/sell}$ is the number of buy/sell trades and $vol_{k,i,d}^{buy/sell}$ gives the trade size of the k-th buy/sell trade.

We implement the size-adapted average bid-ask spread measure using the iterative two-stage weighted regression described in Section 2.2.2. We have to adjust this procedure to account for the fact that there is only one observation per day. The details on the adjustments are described in Appendix A.3. The data requirements to estimate an individual scaling factor remain the same as for the unadapted average bid-ask spread measure. Thus, a monthly size-adapted bid-ask spread measure can be calculated if we observe sell and buy prices for at least one day.

The resulting transaction cost function is again depicted in Figure 2.2. Both size-adapted measures share the decreasing transaction cost and size relation and the slight increase at the right tail. They are not fully identical, though, which is caused by the different data requirements of the standard measures. Whereas the Schultz (2001) measure requires a daily consensus price, the average bid-ask spread measure can be calculated only

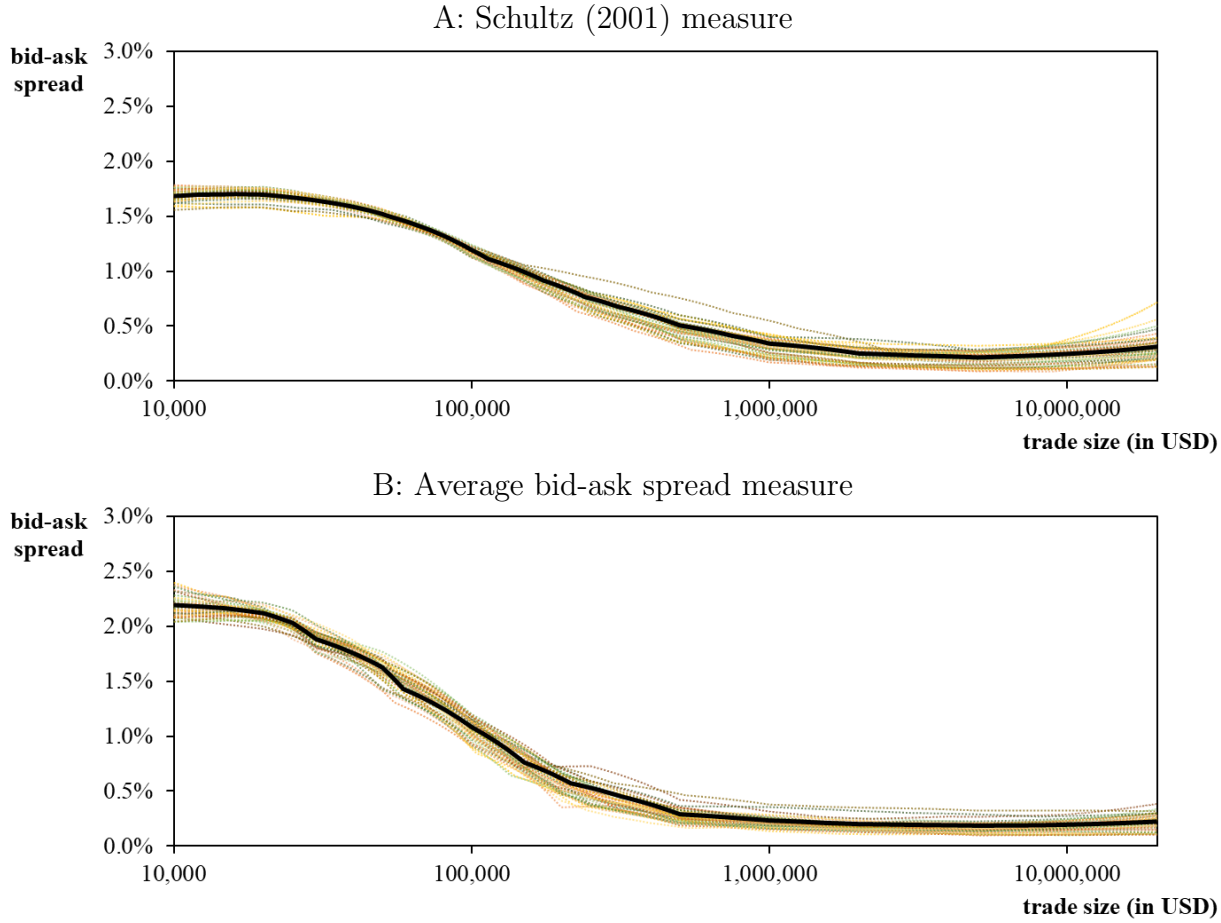


Figure 2.3: **Time variation of transaction cost functions**
Panels A and B depict quarterly transaction cost functions for the Schultz (2001) and average bid-ask spread measure. Transaction cost functions are estimated with the two-step measurement approach described in Sections 2.2.2 and 2.2.3. We use either all observations (black solid line) or all observations in a quarter (dashed lines) for the estimation. Quarterly transaction cost functions are estimated for the time period from October 2004 to December 2014. To ensure comparability, we scale the quarterly transaction cost functions to the average level of the function that is based on all observations.

when data on buy and sell trades are available on the same day.

Having introduced both size-adapted measures, we can now verify that the general transaction cost function does not change much over time and that small deviations have no significant impact on the scaling factors. First, in Figure 2.3, we plot a (normalized) function for each quarter during our observation period. For both liquidity measures, the pattern of a strongly decreasing function and a slight increase at the right tail is stable over time. All quarterly functions are located within a narrow range. Next, we replace our estimated transaction cost functions for both liquidity measures with an interpolated function based on average trading costs for different trade sizes from Edwards, Harris, and Piwowar (2007).[13] Calculating the monthly scaling factors relative to this function, we observe for both liquidity measures extremely strong average cross-sectional correlations of 0.97 and 0.98 between the scaling factors of the original and the interpolated curve for the Schultz (2001) and the average bid-ask spread measures, respectively. Consistent with these high correlations, we find that our results in the following sections remain unchanged when employing the Edwards, Harris, and Piwowar (2007) transaction cost function.

Finally, Figure 2.4 shows average liquidity for all bonds calculated with either the two standard measures or the two size-adapted measures for the time period of October 2004 to December 2014. All measures show a strong increase of transaction costs during the financial crisis. Comparing standard and size-adapted measures, the increase is stronger for the size-adapted measures, pointing to a possible underestimation of liquidity deterioration by the standard measures. In contrast, standard measures overestimate the post-crisis market liquidity compared to the pre-crisis level (see, e.g., Choi and Huh (2019)). While the standard measures indicate a clear improvement in market liquidity, the size-adapted measures point to a rather similar level of liquidity pre and post crisis. Finally, the two size-adapted measures are very consistent and move very closely together.

2.3 Precision of Liquidity Measurement

From a conceptual perspective, the main difference between standard liquidity measures and their size-adapted counterparts is the dependence on the underlying trade-size pattern. As discussed in the introduction, changes in the average trade size must not necessarily relate to changes in a bond's "true" liquidity but are mechanically linked to changes in

[13] See their Table IV for transaction costs of trade sizes between \$5,000 and \$10 million. We use a flat extrapolation for trades below and above these thresholds.

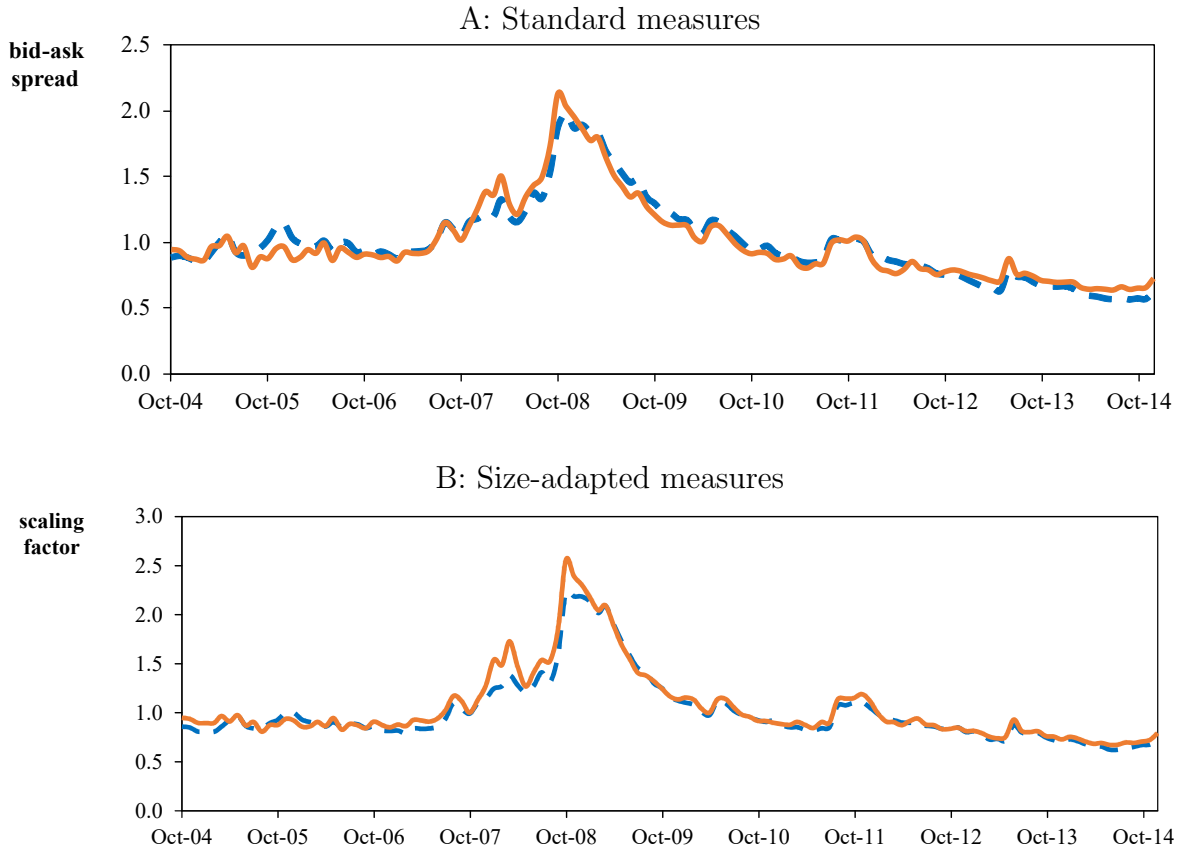


Figure 2.4: **Time series of U.S. corporate bond transaction costs**
The figure depicts the time series of average liquidity across all bonds for the Schultz (2001) measure (orange solid lines) and the average bid-ask spread measure (blue dashed lines). In Panel A, we employ the standard version of the two liquidity measures and in Panel B their size-adapted counterpart of Sections 2.2.2 and 2.2.3. The time period spans from October 2004 to December 2014. Both size-adapted measures can be interpreted as a scaling factor that, together with the transaction cost functions in Figure 2.2, can be used to calculate transaction costs for arbitrary trade sizes. For ease of comparability, we scale the standard measures such that the time-series average equals 1.

the standard liquidity measure.

To examine the mechanical relation between trade size and liquidity measures, we calculate average monthly cross-sectional correlations of the liquidity measures with the bond's (logarithmic) average trade size. We further calculate correlations with other bond characteristics such as rating, age, duration, and amount outstanding.[14]

The results for the standard and size-adapted versions of the Schultz (2001) and the average bid-ask spread measures are presented in Panel A and B of Table 2.1. As expected, we see that standard and size-adapted measures show a different relation to the average trade size. While the size-adapted Schultz (2001) measure only exhibits a rather low correlation of -0.12 and the size-adapted average bid-ask spread shows virtually no correlation, both standard measures are strongly and significantly correlated to trade size with correlations of -0.40 and -0.48. Thus, consistent with the intuition illustrated in Figure 1, a smaller average trade size leads standard measures to report higher transaction costs. In contrast, we find similar correlations of standard and size-adapted measures with rating, age, duration, and amount outstanding. Interestingly, we see high correlations of 0.68 and 0.85 between the standard measures and their size-adapted versions showing a shared common liquidity component.

In the remainder of this section, we show that the strong dependence of the standard measures on average trade size leads to more noisy and, under particular circumstances, biased measures of liquidity.

2.3.1 Unsystematic Measurement Error

Comparing the precision of liquidity measures is difficult because a bond's true liquidity is not observable. We therefore propose two alternative approaches. First, we construct an artificial benchmark from highly traded bonds because their liquidity can be measured with the highest possible precision. Second, we employ an indirect approach exploiting the liquidity premium embedded in corporate bond yields. Because changes in a bond's (true) liquidity should be reflected in changes in the yield spread, we expect measures with lower estimation noise to be better able to explain yield spread shifts.

[14]We measure a bond's rating as the average numerical rating across the three rating agencies S&P, Fitch, and Moody's (S&P and Fitch AAA=1, AA+=2, ... and for Moody's Aaa=1, Aa1=2, ...). We test the average cross-sectional correlations for significance by first transforming the monthly correlations with Fisher's Z and then running a t-test on the time series of the transformed values with Newey and West (1987) correction of six lags.

Table 2.1: **Average cross-sectional correlations**
This table reports average cross-sectional correlations for the standard and size-adapted liquidity measures with average trade size and various bond characteristics. We calculate the logarithm of a bond's monthly average trade size from TRACE as well as its age and duration. The credit rating is measured as the average across numerical ratings from S&P, Moody's, and Fitch (AAA=1, AA+=2, ...). The history of the outstanding amount is from Reuters Eikon. In Panel A, we employ the Schultz (2001) measures of Section 2.2.2 and in Panel B the average bid-ask spread measures of Section 2.2.3. We test the average cross-sectional correlations for significance by running a t-test with Newey and West (1987) correction of six lags on their Fisher's Z transformed time series. ** and * represent statistical significance at the 1% and 5% level.

Panel A: Schultz (2001) measure						
	Average trade size	Rating	Age	Duration	Amount outstanding	Size-adapted measure
Standard measure	-0.40**	0.17**	0.19**	0.32**	-0.28**	0.85**
Size-adapted measure	-0.12**	0.21**	0.12**	0.41**	-0.14**	
Panel B: Average bid-ask spread measure						
	Average trade size	Rating	Age	Duration	Amount outstanding	Size-adapted measure
Standard measure	-0.48**	0.06*	0.13**	0.35**	-0.38**	0.68**
Size-adapted measure	-0.01	0.20**	0.07**	0.42**	-0.11**	

In the first, direct approach, we randomly delete trades of the synthetic benchmark bonds, mirroring the trading pattern of scarcely and moderately traded bonds. We then calculate the liquidity measures based only on the information from the remaining trades. Finally, we approximate the estimation noise as the difference of the liquidity measures calculated with the full and partial information set. We implement this exercise in a bootstrap-like manner, randomly assigning the trading patterns of bonds picked from one of four different trading activity categories (see Appendix A.4 for a more detailed description of the bootstrapping). The four trading activity categories are less than 5 trades per month, 5 to 9 trades, 10 to 19 trades, and 20 to 30 trades. They cover about 50% of the observation months in our sample.

The results of the bootstrapping are presented in Panel A of Table 2.2 for the Schultz (2001) measure and in Panel B for the average bid-ask spread measure. Consistent with our expectation, we find that the measurement errors decrease monotonously for trading categories with more trades. For the standard liquidity measures, the average root-mean-squared percentage error (RMSPE) ranges between 0.586 and 0.858 for 5 to 9 trades and between 0.335 and 0.543 for 20 to 30 trades. For less than 5 trades, it is particularly difficult to capture liquidity. As a result, the measurement errors are with 0.838 and 0.989 for the standard Schultz (2001) and average bid-ask spread measure relatively high. When measuring liquidity with our size-adapted measures, we find that the errors are smaller compared to their standard counterparts. On average, the RMSPEs are roughly 17% to 30% lower when the liquidity measures are calculated based on 5 to 30 trades. When trading is extremely scarce with less than 5 trades available, we still observe between 8% to 9% lower measurement errors with our size-adapted measures. Summarizing, our new size-invariant measurement approach decreases estimation noise consistently for scarcely and for moderately traded bonds.

For the second, indirect approach, we exploit Amihud and Mendelson's (1986) hypothesis that an illiquid asset commands a higher (expected) return and thus trades at a larger spread between the bond's yield and the Treasury curve. Given idiosyncratic changes in the average trade size, especially for rarely traded bonds, we expect that monthly changes in the standard measures are less informative for the liquidity related changes in a bond's yield spread. To test our hypothesis, we follow closely the approach of Friewald, Jankowitsch, and Subrahmanyam (2012) and perform panel data regressions of yield spreads on transaction cost measures, while controlling for autocorrelation in the spreads and other

Table 2.2: **Estimation noise in scarcely traded bonds**
This table reports estimation errors when mirroring the trading pattern of scarcely and moderately traded bonds on the bond observation months with the 1% highest number of trades, which serve as benchmark. We employ four different trading activity categories ranging from less than 5 trades up to 20 to 30 trades. In each of the 100 runs of the bootstrap-like procedure, bonds of the respective trading activity category are selected and assigned randomly to a heavily traded benchmark bond. We then select the most appropriate trades of the benchmark bond to mirror the trading pattern of the scarcely traded bond and estimate the liquidity measures based on the full and the reduced information set. Lastly, we calculate the root-mean-squared percentage error (RMSPE) as the difference between the two estimates. The results of the average estimation errors for the standard and size-adapted Schultz (2001) measure of Section 2.2.2 are presented in Panel A and for the two versions of the average bid-ask spread measure of Section 2.2.3 in Panel B.

Panel A: Schultz (2001) measure				
	trades			
	<5	5 to 9	10 to 19	20 to 30
Standard liquidity measure	0.989	0.858	0.673	0.543
Size-adapted measure	0.900	0.709	0.532	0.429
$\Delta(RMSPE)$	-9.0%	-17.4%	-21.0%	-21.0%
Observations	3,880			
Panel B: Average bid-ask spread measure				
	trades			
	<5	5 to 9	10 to 19	20 to 30
Standard liquidity measure	0.836	0.586	0.424	0.335
Size-adapted measure	0.768	0.445	0.296	0.246
$\Delta(RMSPE)$	-8.1%	-24.1%	-30.2%	-26.6%
Observations	5,230			

effects:

$$\begin{aligned}\Delta(\text{Yield spread})_{i,t} = \alpha + \beta \cdot \Delta(\text{Yield spread})_{i,t-1} + \gamma \cdot \Delta(\text{Transaction costs})_{i,t} \\ + \delta \cdot \Delta(\text{Controls})_{i,t} + \epsilon_{i,t},\end{aligned} \tag{2.6}$$

where we use either the standard or the size-adapted versions of the Schultz (2001) or the average bid-ask spread measure as proxy for transaction costs. We calculate yield spreads as the difference between the bond's yield and the yield of a (theoretical) risk-free Treasury bond having the same cash flow structure. To obtain a bond's daily yield spread, we calculate the volume-weighted average from all reported trades in TRACE in the bond on the specific day (for details, we refer to Appendix A.1). We then take averages of the daily observations to arrive at the monthly level. In the choice of the control variables, we lean on Friewald, Jankowitsch, and Subrahmanyam (2012). We control for monthly changes in the logarithm of the amount outstanding (our data from Reuters Eikon includes reopenings, repurchases, and other (early) redemptions). Moreover, we employ the logarithm of the average trade size and the number of trades. Finally, we include changes in 21 rating dummies based on the bond's average numerical rating across the three agencies Standard & Poor's, Moody's, and Fitch.[15] Yield spread changes and transaction cost changes are winsorized each month at the 1% and 99% level.

The results of the monthly panel regression for both the Schultz (2001) and the average bid-ask spread measure are presented in Table 2.3. For both measures, we use three different specifications. First, in specifications (1) and (4), we explain yield spread changes solely with their first lag and the control variables. This specification is the baseline from which we can analyze the increase in explanatory power after including a transaction cost measure. In specifications (2) and (5), we include the standard measures, while in (3) and (6), we add the size-adapted versions.

In the specifications without a transaction cost measure, we find a positive but insignificant autocorrelation of yield spread changes.[16] The remaining control variables are broadly consistent with the findings in Friewald, Jankowitsch, and Subrahmanyam (2012).

[15] To calculate the average numerical rating, we transform the ratings to integer numbers (for S&P and Fitch AAA=1, AA+=2, ... and for Moody's Aaa=1, Aa1=2, ...). For $k = 1, \ldots, 21$, we then set the k-th rating dummy to 1, if $k - 0.5 \leq$ average rating $< k + 0.5$.

[16] In contrast to our findings, Friewald, Jankowitsch, and Subrahmanyam (2012) observe a negative autocorrelation of yield spreads which is attributable to microstructure noise. However, our positive autocorrelation coefficient is consistent with Duffee (1998). Note that if we follow Friewald, Jankowitsch, and Subrahmanyam (2012) and switch our observation frequency from monthly to weekly, we also find a negative estimate. Thus, microstructure noise seems to be less important on the monthly level.

Table 2.3: **Yield spread regressions: All trades**

This table reports results for the panel regression model (2.6) explaining monthly yield spread changes with monthly changes in standard or size-adapted transaction cost measures. The control variables are trade size, number of trades, and amount outstanding. We further employ rating dummies as a proxy for credit risk. Specifications (1) and (4) are the baseline regressions that include only lagged yield spread changes and the control variables. In (2) and (5), we add standard transaction cost measures, while in (3) and (6), we add the size-adapted versions. We winsorize yield spread and transaction cost changes at the 1% and 99% level. Standard errors are clustered monthly and by bond. We test for differences between out-of-sample mean squared errors (MSE) using the test statistic proposed by Harvey, Leybourne, and Newbold (1997). Comparing R^2_{adj} or MSE, we first compare a specification to the baseline and second to its preceding specification. The t-statistics are given in parentheses. ** and * represent statistical significance at the 1% and 5% level.

	Schultz (2001)			Average bid-ask spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0229	0.0234	0.0236	0.0422	0.0441	0.0442
	(0.80)	(0.82)	(0.83)	(1.08)	(1.16)	(1.16)
Δ(Yield spread)$_{i,t-1}$	0.0960	0.0955	0.0956	0.0906	0.0896	0.0893
	(1.43)	(1.42)	(1.42)	(0.95)	(0.96)	(0.95)
Δ(Standard measure)$_{i,t}$		5.5337**			9.6839**	
		(6.95)			(5.18)	
Δ(Adapted measure)$_{i,t}$			0.0831**			0.1679**
			(5.58)			(4.88)
Δ(Trade size)$_{i,t}$	-0.0132**	-0.0065	-0.0149**	-0.0259**	-0.0124*	-0.0353**
	(-4.23)	(-1.79)	(-5.30)	(-8.59)	(-2.55)	(-12.22)
Δ(Trades)$_{i,t}$	0.0055**	0.0052**	0.0052**	0.0076**	0.0071**	0.0071**
	(4.39)	(4.26)	(4.28)	(4.91)	(4.70)	(4.90)
Δ(Amount outstanding)$_{i,t}$	-0.0576	-0.0499	-0.0519	-0.1527	-0.1412	-0.1382
	(-0.82)	(-0.71)	(-0.74)	(-1.90)	(-1.74)	(-1.75)
Δ(Rating dummies)$_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes
R^2_{adj}	0.0776	0.0810	0.0836	0.0636	0.0701	0.0759
$\Delta(R^2_{adj})$		4.4%	7.7%/3.2%		10.2%	19.3%/8.3%
MSE	0.909	0.905	0.902	1.683	1.670	1.656
Δ(MSE)		-0.4%**	-0.8%**/-0.3%**		-0.77%**	-1.60%**/-0.84%**
		(6.24)	(7.35)/(4.96)		(9.39)	(13.27)/(9.37)
Observations		327,251			454,461	

If we add the standard bond liquidity measures to the regression, we find a highly significant coefficient. Regarding the explainable part of yield spread changes, adding a standard transaction cost measure leads to a relative improvement of the R^2_{adj} compared to the baseline regression of 4.4% in case of the Schultz (2001) measure and of 10.2% in case of the average bid-ask spread. The absolute improvements in the R^2_{adj} range between 0.0034 and 0.0065 and are rather low, which is usually the case in this type of regression setup. Relative and absolute improvements are quantitatively comparable to the ones in Friewald, Jankowitsch, and Subrahmanyam (2012). These authors add four different liquidity measures at the same time and find a combined relative improvement of 11.7% and an absolute improvement of 0.009. When adding our size-adapted measures instead of the standard versions, we also find a highly significant coefficient. The relative improvement of the R^2_{adj} compared to the baseline regression increases to 7.7% in case of the size-adapted Schultz (2001) measure and to 19.3% in case of the size-adapted average bid-ask spread measure.

We now validate the higher explanatory power of our less noisy size-adapted measures out-of-sample. To this end, we estimate the panel model (2.6) using a backward-looking rolling window of 24 months (requiring at least 12 months) and compare the implied with the actual yield spread changes for the following month. To test if changes in the mean squared error (MSE) are significant, we employ Diebold and Mariano (1995) tests in the spirit of Harvey, Leybourne, and Newbold (1997).[17] Consistent with our in-sample findings, we observe at the bottom of Table 2.3 a monotonically and significantly decreasing out-of-sample MSE when adding a transaction cost measure to the baseline model. The additional improvements when incorporating a size-adapted liquidity measure are more than twice as large and highly significant as well.

In the robustness section 2.5.1, we run yield-spread regressions using the approaches from the literature to account for the larger transaction costs of small trades. It turns out that employing only large trades with volumes of at least \$100,000 or weighting by trade size cannot improve the explanatory power of transaction cost changes in such regressions compared to the standard, equally weighted liquidity measures. This result supports the findings from Appendix A.2 that small trades contain valuable information. We conduct several other unreported robustness tests (available on request). First, we show that the

[17] Since we use a rolling window to estimate parameters, the Diebold and Mariano (1995) test is identical to Giacomini and White's (2006) (unconditional) test that accounts for uncertainty in the parameter estimation (see, e.g., Giacomini and Rossi, 2010). Furthermore, in Chapter 4, we design a test procedure specifically for out-of-sample tests in the presence of a large panel. As a robustness check, we repeat the analysis applying this test procedure in Appendix D.

results are not subject to a look-ahead bias by using only data from the last quarter of 2004 (and not full sample information) to estimate the transaction cost function. Second, we show that the different improvements between the Schultz (2001) and the average bid-ask spread measure are partly due to their different data requirements. Third, we find similar additional explanatory power for our size-adapted measures when we employ a bond's 1-year probability of default instead of rating dummies to proxy for credit risk. Lastly, we show that our findings do not change if we exclude prearranged roundtrip transactions in the spirit of Choi and Huh (2019) that do not provide immediacy.

Summarizing, we first show directly that our size-invariant approach offers a higher measurement precision as it is less prone to estimation noise in situations when only few observations are available. Second, we employ an indirect approach showing that the more precise size-adapted measures consistently increase the explainable part of yield spread changes compared to standard liquidity measures.

2.3.2 Liquidity Measurement Bias During Fire Sales

In the previous section, we showed that our size-adapted liquidity measures are more precise due to a lower unsystematic measurement error. This advantage is relevant when bonds trade scarcely and their changes in the trade-size pattern are mostly idiosyncratic. However, during stress events when large investors are forced to sell their positions, the associated changes in average trade sizes become systematic and may lead to a measurement bias in the standard liquidity measures.

To test this hypothesis, we use rating downgrades from investment grade to junk status as a natural experiment. Ambrose, Cai, and Helwege (2008) and Ellul, Jotikasthira, and Lundblad (2011) argue that insurance companies are forced to sell their bond positions in such situations due to regulatory constraints. Bao, O'Hara, and Zhou (2018) find that selling usually happens right after the downgrade event. Thus, the deterioration of the downgraded bond's liquidity should be accompanied by a strong increase in total trading volume along with increased average trade sizes for sell trades. As larger trades pay lower transaction costs, measuring liquidity with a standard measure could lead to a downward bias and veil the increased illiquidity to an observer.

We follow Bao, O'Hara, and Zhou (2018) and define the downgrading day as the first day on which one of the three major rating agencies S&P, Moody's, or Fitch downgrades an investment grade bond to a speculative grade rating of BB+/Ba1 or lower. We calculate

the trading volume and the average trade size in the downgraded bond in the month after the downgrade (including the downgrading day). We then compare the trading activity in the downgrading month to the trading activity in the 30 days preceding the downgrade. Panel A of Table 2.4 presents the changes in trading activity. We see a significant average increase in the monthly trading volume of nearly 700%. Also, the average trade size increases significantly by more than 200%. To test whether this change is the result of the downgrade, we follow Bao, O'Hara, and Zhou (2018) in using bonds that are rated BB+ to BB- (Ba1 to Ba3) in the downgrading month and the month before the downgrade as control group. As expected, there is no significant change in the trading volume or the average trade size for the average bond in the control group. Finally, to see if rating downgrades are associated with an increased selling pressure, we follow Bai and Collin-Dufresne (2011) and use the percentage change in selling volume as a proxy of bond selling pressure. Indeed, we observe a significant increase in selling volume of more than 800% for the downgraded bonds and an insignificant change for the peer bonds. Interestingly, the increased selling volume is accompanied by a significant increase in the average trade size.

In Panel B, we look at the changes in liquidity during downgrade-induced fire sales when measuring the liquidity with either the standard or the size-adapted measures of Sections 2.2.2 and 2.2.3. As Anderson and Stulz (2017) argue, not all downgraded bonds suffer from selling pressure as they are not necessarily held by investors who are required to sell. We therefore restrict our sample to bonds with likely forced sells, for which we require that selling volume in the downgrade month is greater or equal to, first, buying volume in the downgrade month and, second, selling volume in the month before. Third, we require that selling turnover, i.e., selling volume normalized with the outstanding amount, in the downgrade month is larger than median selling turnover in our full sample. We then calculate the two standard liquidity measures and their size-adapted counterparts for the month before and after the downgrade event. For comparability, we scale all measures with their full sample mean. For both standard liquidity measures, the average difference between the month before and after the downgrade is not significantly different from zero. In contrast, our new size-adapted measures show, as expected, a significant increase in transaction costs during the fire sale. Contrary to the expectation that transaction costs increase after the stress event, the difference between the transaction cost change of the downgraded bonds and the bonds in the peer group is (insignificantly) negative when we employ standard measures. For both size-adapted measures, the same difference is significantly positive, confirming hypotheses from theoretical models (see, e.g., Brunner-

Table 2.4: **Liquidity measurement bias during fire sales**
This table reports changes in trading activity and in liquidity between the month before a bond is downgraded from investment grade to junk status and the month starting with the downgrade date. The downgrade date is set as the day for which the first acting rating agency, S&P, Moody's, or Fitch, rates a previously investment-grade rated bond as BB+/Ba1 or lower. For each downgrading event, we construct a control group of bonds that have been rated BB/Ba during the month before and the month starting with the downgrade event. In Panel A, we show average percentage changes in total trading volume and in average trade size based on all trades and based only on customer sell trades for the downgraded bonds and for the peer bonds in the control group. In Panel B, we report changes in liquidity between the month before the downgrade and the downgrade month for bonds which are likely to experience selling pressure (see Section 2.3.2 for details). We measure liquidity using the standard or size-adapted Schultz (2001) and average bid-ask spread measures. We also report the difference between the changes of the downgraded bonds and the control group, as well as differences between the standard and the size-adapted measures (diff-in-diff). To ensure comparability, we scale standard and size-adapted measures by their sample mean. We winsorize changes in trading activity and liquidity at the 1% and 99% level. Standard errors are clustered by firm and t-statistics are given in parentheses. ** and * represent statistical significance at the 1% and 5% level.

Panel A: Changes in trading activity after a rating downgrade

	All trades		Sell trades	
	Trading volume (in %)	Avg. trade size (in %)	Trading volume (in %)	Avg. trade size (in %)
Downgraded	699.68**	237.72**	818.43**	233.85**
	(4.99)	(4.79)	(4.40)	(4.91)
Peer bonds	-4.30	0.05	-3.98	-2.47
	(-0.92)	(0.02)	(-0.75)	(-0.88)
Observations	2,386			

Panel B: Changes in liquidity during a fire sale

	Schultz (2001)		Average bid-ask spread	
	Standard	Size-adapted	Standard	Size-adapted
Downgraded &	0.014	0.189*	-0.073	0.165*
selling pressure	(0.22)	(2.33)	(-1.33)	(2.21)
Peer bonds	0.021*	0.041**	-0.042**	-0.023
	(2.31)	(2.77)	(-3.19)	(-1.84)
Difference to peer	-0.007	0.148*	-0.031	0.188*
	(-0.10)	(2.04)	(-0.49)	(2.58)
Diff-in-diff	0.155**		0.219**	
	(2.73)		(2.78)	
Observations	336		399	

meier and Pedersen, 2009). Lastly, the difference between the size-adapted measure and the standard measure (difference-in-difference) is significant for both the average bid-ask spread measure as well as the Schultz (2001) measure.

In summary, during stress events such as fire sales, the shift to a trading pattern with larger trades leads to a downward bias in standard measures of liquidity and potentially misleading conclusions regarding the true state of liquidity. As our size-adapted measures explicitly take the trade-size pattern into account, they are not subject to this bias. They could therefore help to reconcile the conflicting results whether post-crisis liquidity is lower during stress events due to stricter regulation. Whereas Anderson and Stulz (2017) do not find evidence for worse liquidity post-crisis using a battery of different liquidity measures, Dick-Nielsen and Rossi (2019) and Bao, O'Hara, and Zhou (2018) find that liquidity after stress events has decreased. Bao, O'Hara, and Zhou (2018) base their evidence mainly on the Amihud (2002) measure, which is closely related to volatility and, for bond markets, only relatively weakly correlated with transaction cost measures (see Schestag, Schuster, and Uhrig-Homburg, 2016). Dick-Nielsen and Rossi (2019) develop their own measure based on the returns of liquidity providing dealers.

2.4 The Pricing of Corporate Bond Liquidity

In this section, we analyze the effects of corporate bond liquidity and corporate bond market liquidity risk on expected bond returns using our new measurement approach. Previous studies find that investors are compensated for the individual level of a bond's transaction costs (see, e.g., Bao, Pan, and Wang, 2011). A second strand of the literature shows that investors require a premium for a bond's sensitivity to corporate bond market illiquidity shocks (see, e.g., Lin, Wang, and Wu, 2011). Bongaerts, de Jong, and Driessen (2017) analyze both effects jointly and find that only the level of liquidity but not corporate bond market liquidity risk bears a premium. We aim to reexamine this question in this section, exploiting the higher precision of our size-invariant measurement approach. From an econometric perspective, a precise measurement of liquidity is important to disentangle the effects of liquidity level and risk.

Because asset pricing results can depend on the construction of test assets (see, e.g., Lewellen, Nagel, and Shanken, 2010 and Ang, Liu, and Schwarz, 2020) and the estimation of betas, we introduce four different specifications: Three portfolio sorts using different criteria and one approach estimating the model on an individual bond basis. This exercise

shows that the results using standard measures of liquidity are not robust to the chosen approach. In contrast, when we use our more precise size-adapted measures, we can consistently show that both individual bond liquidity and corporate bond market liquidity risk are priced.

2.4.1 Asset Pricing Model

Acharya and Pedersen (2005) argue that liquidity level and risk should be analyzed jointly due to the correlation of the two variables. In this spirit, Bongaerts, de Jong, and Driessen (2017) propose an asset pricing model, in which both the exposure to systematic bond market illiquidity shocks as well as the individual liquidity level can have an impact on expected bond returns. Additionally, they consider spill-over effects from the equity market. We follow these authors and estimate the following asset pricing model. In the first step, we regress the monthly realized excess returns $r_{j,t}$ on the risk factors equity market return EQ_t, equity market illiquidity shocks $EQLIQ_t$, and corporate bond market illiquidity shocks $CBLIQ_t$,

$$r_{j,t} = \beta_j^0 + \beta_j^{\text{EQ}} \cdot EQ_t + \beta_j^{\text{EQLIQ}} \cdot EQLIQ_t + \beta_j^{\text{CBLIQ}} \cdot CBLIQ_t + \epsilon_{j,t}, \tag{2.7}$$

for each portfolio j. In the second step, we regress an estimate of monthly expected excess returns $E[r_{j,t+1}]$ (see below) on the cross-section of risk sensitivities, which are estimated on a rolling basis in the first step, and on the portfolios' transaction cost estimates $c_{j,t}$

$$E[r_{j,t+1}] = \lambda_0 + \lambda_{\text{EQ}} \cdot \beta_{j,t}^{\text{EQ}} + \lambda_{\text{EQLIQ}} \cdot \beta_{j,t}^{\text{EQLIQ}} + \lambda_{\text{CBLIQ}} \cdot \beta_{j,t}^{\text{CBLIQ}} + \lambda_{\text{c}} \cdot c_{j,t} + \alpha_j. \tag{2.8}$$

We obtain market prices of the risk factors and the impact of liquidity level as the time series averages of the monthly cross-sectional estimates for which we calculate Fama-MacBeth standard errors.

We expect bond returns to decrease in times of equity market turmoils leading to positive estimates for β_j^{EQ} and a positive market risk premium λ_{EQ}. As increases in bond market illiquidity should lead to decreasing bond returns, we expect negative estimates for β_j^{CBLIQ} in Equation (2.7). Further, if investors require a compensation for systematic corporate bond market liquidity risk, more negative betas should lead to higher expected returns, implying negative estimates for λ_{CBLIQ}. Given several possible mechanisms for the influence of the equity market illiquidity on bond returns β_j^{EQLIQ}, we have no ex-ante expectation for its sign. In line with Amihud and Mendelson (1986), we expect a positive

λ_c as illiquid bonds should compensate investors with higher returns.

2.4.2 Risk Factors, Returns, and Transaction Costs

For the time series regression (2.7), we calculate a bond's realized return between the last trading days in months t and $t-1$. To obtain excess returns, we deduct the return of a U.S. treasury bond having a maturity equal to the bond's duration.[18] For the equity risk factors, we employ excess returns of the S&P 500 as equity market returns. We characterize equity market liquidity using the Amihud (2002) liquidity measure and calculate its equally weighted mean from all shares having a share code of 10 or 11 in CRSP.[19] As it is common practice, we exclude observations of extremely illiquid stocks, i.e., days without trading and months having less than three days with positive trading volume. Further, we exclude shares traded at NASDAQ[20] and winsorize the monthly cross-section of individual liquidity measures at the 5% and 95% level. Given the monthly aggregate Amihud (2002) liquidity measure, we identify illiquidity shocks as residuals of the autoregressive model proposed in Acharya and Pedersen (2005). Finally, we measure corporate bond market liquidity as the average across all portfolios using one of the measures introduced in Section 2.2. Bond market illiquidity shocks are then defined as the residuals of an autoregressive process with two lags. For ease of comparability, we scale all risk factor innovations to have the same standard deviation as the equity market excess returns.

In the cross-sectional regression (2.8), we use the forward-looking expected excess return measure of Bongaerts, de Jong, and Driessen (2017). They argue that the common approach, using realized returns as a proxy for expected returns, leads to extremely noisy estimates. We follow them and approximate forward-looking expected excess returns using the bond's yield corrected for the expected costs of default. This leads to

$$E[r_{i,t+1}] = (1 + y_{i,t}) \cdot (1 - L \cdot \pi_{i,t})^{1/T_{i,t}} - (1 + y_{i,t}^{\text{risk-free}}), \tag{2.9}$$

where we approximate bond i at time t with a zero coupon bond having a maturity equal

[18] We use updated data from Gürkaynak, Sack, and Wright (2007) available from the Federal Reserve to calculate Treasury prices and returns.

[19] In a recent study, Lou and Shu (2017) show that the pricing of the Amihud (2002) price impact measure in the stock market is not driven by price impact but rather by its trading volume component. Therefore, we repeat our analyses in Section 2.4.4 using the high-low measure of Corwin and Schultz (2012). We find that our results are robust to the choice of the stock market liquidity measure.

[20] Pástor and Stambaugh (2003) and Ben-Rephael, Kadan, and Wohl (2015) argue that the inflated volume on NASDAQ would bias the Amihud (2002) measure.

to its duration $T_{i,t}$. Further, we assume default losses to incur only at maturity, leading to an expected return until maturity of $(1 + y_{i,t})^{T_{i,t}} \cdot (1 - L \cdot \pi_{i,t})$, where $y_{i,t}$ is the yield of the bond, L gives the loss given default, and $\pi_{i,t}$ is the cumulative probability of default (PD) over the bond's remaining life. Lastly, we annualize the expected return and deduct the yield of a risk-free U.S. Treasury bond having the same duration.

To calculate the cost of default, we assume a constant loss given default of 60%. We use company-specific PDs from the Risk Management Institute (RMI) of the University of Singapore (see Appendix A.1 for details on the matching of corporate bond data and company-specific PDs). RMI publishes PDs for over 66,000 publicly traded companies based on the forward intensity model of Duan, Sun, and Wang (2012). For more than 33,000 companies, these probabilities are calculated on a daily basis for a large spectrum of maturities.[21] RMI provides cumulative PDs for the maturities 1 month, 3 months, 6 months, 1 year, 2 years, 3 years, and 5 years. We use them to calculate (annualized) conditional PDs for all possible periods (i.e., from 0 to 1 months, from 1 to 3 months, ...). Assuming a flat curve beyond 5 years, we can calculate the cumulative PD $\pi_{i,t}$ corresponding to the bond's duration $T_{i,t}$.

Finally, we calculate a bond's monthly expected excess return $E[r_{i,t+1}]$ as the volume-weighted average from all trades in month t. Regarding bond i's transaction costs $c_{i,t}$, we use one of the standard or size-adapted liquidity measure from Section 2.2.[22] We aggregate returns and transaction costs to the portfolio level by calculating their equally weighted mean.[23]

[21] See NUS-RMI (2016) for more details on the methodology. For a comparison of RMI's PDs with Moody's expected default frequency (EDF) measure, see Berndt (2015).

[22] To ensure a tradeable strategy and to address concerns of a look-ahead bias, we estimate the market-wide transaction cost function for our size-adapted measures based solely on the last quarter of 2004. The liquidity measures based on this quarterly cost function exhibit a strong average cross-sectional correlation to the full sample measures of 0.974 and 0.986 for the Schultz (2001) and the average bid-ask spread measure, respectively. As a result, and consistent with the discussion in Section 2.2.3, the slight variations in the transaction cost function have no major impact on the estimated scaling factors and we verify that they do not alter our results in Section 2.4.4. Further, for comparability between the Schultz (2001) and average bid-ask measure, we use only observations for which we can calculate both measures.

[23] We again winsorize realized and expected excess returns as well as transaction costs each month at the 1% and 99% level.

2.4.3 Four Different Specifications for the Construction of Test Assets and the Estimation of Betas

Our first specification follows Daniel, Titman, and Wei (2001) and Bongaerts, de Jong, and Driessen (2017) in their recommendation to sort by characteristics and factor betas to guarantee sufficient variation across liquidity and liquidity beta in the portfolios. Thus, we form triple-sorted portfolios based on the previous quarters' credit quality, liquidity level, and liquidity beta (see Appendix A.5 for details on the calculation of individual liquidity betas). For the first stage, we proxy a bond's credit quality with its average rating and assign the bonds to three distinct portfolios (terciles). In the second stage, we sort the bonds of each portfolio into liquidity terciles, which we approximate with their amount outstanding. For the last stage, we further divide the bonds of the preceding 9 portfolios into liquidity beta quintiles. In total this sorting leads to 45 portfolios for the Fama-MacBeth procedure. We estimate the betas of the time series regression (2.7) using a rolling window of 24 backward-looking monthly observations (requiring at least 12 observations).

Given the contradictory findings of the literature so far, our second specification is designed to be closely comparable to Bongaerts, de Jong, and Driessen (2017). Their approach differs from our first specification in two distinctive features. First, Bongaerts, de Jong, and Driessen (2017) increase the number of portfolios and allow that a bond is assigned simultaneously to up to six portfolios. Second, to increase precision, betas are estimated using a forward-looking kernel, which makes it impossible to collect the risk premia with a tradeable strategy. Regarding the first feature, the authors use multiple proxies for a bond's credit quality and liquidity for the first and second stage of the triple sorting. The sort on credit quality is either done using a bond's average rating (AAA-A, BBB, and BB-CCC) or its RMI 1-year cumulative PD (terciles). For the second sorting stage, a bond is classified as liquid or illiquid using either its amount outstanding (median), age (median), or number of trades (70% percentile). Using the different proxies for credit risk and liquidity, a bond is simultaneously assigned to 6 portfolios. Given the 6 portfolios of the two sorting steps, the number of portfolios then increases to 36. These portfolios are finally split into a high-liquidity-beta and low-liquidity-beta category, leading to a total number of 72 portfolios. Regarding the second feature, they run the time series regression (2.7) using a two-sided (triangular) kernel for the rolling beta estimation. This tent-shaped kernel is both forward- and backward-looking, with linearly decreasing weights up to a maximum distance of 12 months relative to the current observation date. We require

12 backward-looking observations and use a truncated form of the kernel for dates with less than 12 future observations.

Further, we account for the suggestion in Lewellen, Nagel, and Shanken (2010) that asset pricing tests should include additional portfolio sorts on other characteristics. In this spirit, we introduce a third approach in which we sort bonds on their industry affiliation in the previous quarter. We use the 4-digits GICS code of the bond's issuer to ensure that each portfolio includes a sufficient number of bonds. Finally, we test our asset pricing model for individual bonds, given the argument in Ang, Liu, and Schwarz (2020) that portfolio sorts render the estimation of risk premia inefficient due to a loss in cross-sectional variation of the factor loadings. In both settings, we estimate the time series regression (2.7) using a backward-looking rolling window of 24 months. Because many bonds do not trade consecutively, we require for the individual bonds at least 12 observations.

2.4.4 Results

The results of the Fama-MacBeth regressions for all four specifications are reported in Table 2.5. We start with examining corporate bond liquidity effects that are identified by using the standard bond liquidity measures. The first and third column of Panel A display results based on the first setting – the non-overlapping triple sort. We find a significantly negative λ_{CBLIQ} in both specifications, which, in combination with the negative average β^{CBLIQ}, implies a positive corporate bond market liquidity risk premium.[24] Moreover, we find a positive liquidity level premium λ_{c}. These results contradict the significant level and insignificant corporate bond market liquidity risk premium in Bongaerts, de Jong, and Driessen (2017). Therefore, we run the Fama-MacBeth regressions again using their measure of transaction costs based on a repeat-sales method.[25] The results in specification (5) show a significant bond market liquidity risk but insignificant level premium. Thus, common measures show either a pricing of both liquidity level and risk or solely a pricing of market liquidity risk.

Given the contradictory findings in the first setting, we proceed to the results of the second setting in Panel B – the original approach of Bongaerts, de Jong, and Driessen (2017) using overlapping portfolios. For both standard measures in specifications (1) and (3), we now find a premium for corporate bond market liquidity risk but no premium for

[24]For summary and correlation statistics on expected excess returns, betas, and liquidity level, we refer to Appendix A.6.

[25]For details, see Section 1.3 of Bongaerts, de Jong, and Driessen (2017).

Table 2.5: **Fama-MacBeth regression**
This table reports time series averages of the monthly results from the cross-sectional Fama-MacBeth regression (2.8). We estimate premia for equity market risk (λ_{EQ}), equity market liquidity risk (λ_{EQLIQ}), corporate bond market liquidity risk (λ_{CBLIQ}), and individual bond liquidity (λ_c). Specifications (1) and (3) use the standard measures, whereas (2) and (4) employ their size-adapted counterparts. In specifications (5) and (6), transaction costs are calculated with a standard and size-adapted repeat-sales method. Results in Panels A and B are based on the non-overlapping and overlapping triple sorts. Panels C and D present the results for the industry sort and the individual bonds (for details see Section 2.4.3). The Fama-MacBeth t-statistics are calculated based on Newey and West (1987) standard errors with six lags and are given in parentheses. ** and * represent statistical significance at the 1% and 5% level.

Panel A: Non-overlapping triple sort

	Schultz (2001)		Average bid-ask spread		Repeat-sales	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0036	-0.0010	0.0042	0.0001	0.0058	0.0015
	(1.60)	(-0.76)	(1.36)	(0.08)	(1.68)	(0.58)
λ_{EQ}	0.0583**	0.0538**	0.0594**	0.0530**	0.0613**	0.0606**
	(4.75)	(5.39)	(4.89)	(5.23)	(5.05)	(5.06)
λ_{EQLIQ}	-0.0324**	-0.0211**	-0.0297**	-0.0264**	-0.0324**	-0.0280**
	(-3.54)	(-3.14)	(-3.34)	(-4.26)	(-3.75)	(-3.38)
$\lambda_{\text{CBLIQ}}^{\text{Standard}}$	-0.0393**		-0.0447**		-0.0309**	
	(-5.71)		(-8.07)		(-3.62)	
$\lambda_{\text{CBLIQ}}^{\text{Size-adapted}}$		-0.0493**		-0.0449**		-0.0388**
		(-3.69)		(-4.24)		(-4.56)
$\lambda_{\text{c}}^{\text{Standard}}$	0.4315**		0.3601*		0.5577	
	(3.13)		(2.54)		(1.74)	
$\lambda_{\text{c}}^{\text{Size-adapted}}$		0.0084**		0.0076**		0.0085**
		(4.48)		(4.92)		(4.08)
R^2_{adj}	69.8%	70.4%	70.8%	70.6%	72.1%	71.6%

Panel B: Overlapping triple sort

	Schultz (2001)		Average bid-ask spread		Repeat-sales	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0050	-0.0022	0.0059	0.0001	0.0041	-0.0040
	(1.40)	(-0.72)	(1.56)	(0.04)	(1.37)	(-1.05)
λ_{EQ}	0.0625**	0.0552**	0.0651**	0.0670**	0.0665**	0.0552**
	(3.60)	(3.21)	(3.92)	(5.07)	(4.69)	(3.70)
λ_{EQLIQ}	-0.0365**	-0.0444**	-0.0290**	-0.0338**	-0.0307**	-0.0294**
	(-3.06)	(-2.67)	(-2.69)	(-2.70)	(-3.24)	(-2.99)
$\lambda_{\text{CBLIQ}}^{\text{Standard}}$	-0.0326**		-0.0436**		-0.0166	
	(-2.99)		(-3.28)		(-1.22)	
$\lambda_{\text{CBLIQ}}^{\text{Size-adapted}}$		-0.0526**		-0.0469*		-0.0447**
		(-2.95)		(-2.55)		(-4.84)
$\lambda_{\text{c}}^{\text{Standard}}$	0.4769		0.4342		1.0603*	
	(1.51)		(1.51)		(2.47)	
$\lambda_{\text{c}}^{\text{Size-adapted}}$		0.0122**		0.0076*		0.0153**
		(3.12)		(2.42)		(3.95)
R^2_{adj}	82.1%	82.6%	81.4%	80.7%	83.1%	84.2%

Table 2.5 continued

Panel C: Industry sort

	Schultz (2001)		Average bid-ask spread		Repeat-sales	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0048*	-0.0007	0.0086*	0.0000	0.0069*	0.0054
	(2.13)	(-0.36)	(2.01)	(0.01)	(2.39)	(1.47)
λ_{EQ}	0.0480**	0.0413**	0.0495**	0.0447**	0.0470**	0.0414**
	(6.09)	(5.51)	(6.19)	(6.01)	(6.39)	(5.20)
λ_{EQLIQ}	-0.0178**	-0.0152**	-0.0208**	-0.0148**	-0.0184**	-0.0166**
	(-3.31)	(-4.23)	(-3.45)	(-3.68)	(-4.48)	(-4.36)
$\lambda_{\text{CBLIQ}}^{\text{Standard}}$	-0.0404**		-0.0410**		-0.0344**	
	(-4.11)		(-4.38)		(-3.17)	
$\lambda_{\text{CBLIQ}}^{\text{Size-adapted}}$		-0.0348**		-0.0411**		-0.0426**
		(-2.79)		(-3.03)		(-3.05)
$\lambda_{\text{c}}^{\text{Standard}}$	0.4126*		0.2169		0.6426	
	(2.27)		(0.95)		(1.80)	
$\lambda_{\text{c}}^{\text{Size-adapted}}$		0.0112**		0.0099**		0.0063*
		(5.09)		(4.72)		(1.99)
R^2_{adj}	70.9%	72.8%	70.6%	69.3%	71.7%	72.6%

Panel D: Individual bonds

	Schultz (2001)		Average bid-ask spread	
	(1)	(2)	(3)	(4)
Intercept	0.0083**	0.0079**	0.0095**	0.0085**
	(5.85)	(6.11)	(6.10)	(6.92)
λ_{EQ}	0.0314**	0.0301**	0.0313**	0.0306**
	(4.03)	(3.70)	(4.19)	(4.27)
λ_{EQLIQ}	-0.0154**	-0.0153**	-0.0151**	-0.0155**
	(-2.98)	(-2.97)	(-3.15)	(-3.34)
$\lambda_{\text{CBLIQ}}^{\text{Standard}}$	-0.0261**		-0.0270**	
	(-3.21)		(-3.53)	
$\lambda_{\text{CBLIQ}}^{\text{Size-adapted}}$		-0.0221**		-0.0259**
		(-2.70)		(-3.02)
$\lambda_{\text{c}}^{\text{Standard}}$	0.4088**		0.3547**	
	(5.07)		(4.48)	
$\lambda_{\text{c}}^{\text{Size-adapted}}$		0.0061**		0.0047**
		(6.61)		(8.38)
R^2_{adj}	41.3%	39.9%	38.9%	36.5%

liquidity level. In contrast to the previous results but consistent with Bongaerts, de Jong, and Driessen (2017), we find for the repeat-sales measure in (5) a significant level and an insignificant corporate bond market liquidity risk premium. Again, common measures do not yield a consistent pricing pattern.

Finally, as none of the preceding settings provides a stable pricing pattern, we examine bond liquidity effects of the standard measures based on the third and fourth setting. In Panel C, we show the results for the industry sort and in Panel D for the individual bonds.[26] Again, we find no consistent pricing pattern for the three measures. While the Schultz (2001) measure in both settings and the average bid-ask spread in the fourth setting indicate a pricing of both factors, the repeat-sales measure and the average bid-ask spread in the third setting exhibit only a market liquidity risk premium. Summarizing, none of the standard measures is able to identify a stable pricing pattern of corporate bond liquidity level and risk across different approaches to form portfolios and to calculate betas. While five specifications lead to the conclusion of a pricing of both liquidity effects, five support only a significant market liquidity risk premium and one specification provides evidence for just a liquidity level premium.

In contrast, if we measure bond market and individual bond liquidity with the size-adapted measures of Section 2.2, we find a consistent pricing pattern. In all four panels of Table 2.5, specifications (2) and (4) show a significantly negative λ_{CBLIQ}. In combination with the negative β^{CBLIQ}, this implies a significantly positive corporate bond market liquidity risk premium. Moreover, the significantly positive λ_{c} in all panels confirms a positive premium for the level of individual bond illiquidity. To support the finding that the size-adaptation resolves the conflicting results, we further adapt the repeat-sales measure with our new measurement approach.[27] The size-adapted repeat-sales measures in specification (6) consistently show significant premia for both effects. Lastly, the intercepts in all portfolio settings are insignificant as the theory predicts (see Bongaerts, de Jong, and Driessen (2017) for a discussion). For the standard measures, this result holds only for the two triple sorts. In the individual bond setting, potentially due to the noise in the estimated betas, the intercepts are significant in all specifications. Notably, the estimates decline absolutely across all settings when we employ our size-adapted liquidity measures.

In summary, estimating the asset pricing model using our new size-adapted approach always leads to a significant premium for corporate bond market liquidity risk as well as

[26] Note that the repeat-sales measure is designed to calculate portfolio-wide transaction costs and is not applicable to individual bonds.

[27] For details on the adaptation see Appendix A.7.

for the individual liquidity level. It is interesting to note that the individual approach, in contrast to the portfolio settings, comes to the same conclusion, independent of the choice of the liquidity measure. On the one hand, our results thus support the recommendation of Ang, Liu, and Schwarz (2020) to test asset pricing models using individual assets. On the other hand, the scarce trading in the corporate bond market may render asset pricing results based on an individual approach prone to a sample bias and lead to noisy risk premia estimates. First, many bonds, particularly illiquid ones, are excluded when betas are estimated individually. While our portfolio sorts on average incorporate information of roughly 1,500 to 1,600 bonds per month, the individual approach covers only about 1,000 bonds on average. Second, on average, only about 25% of these bonds span the full window of 24 observations when estimating their betas. Consequently, most betas are exposed to strong estimation noise, causing the risk premia estimates (as product of average beta and lambda estimate) to be noisy as well.

Using the results for the size-adapted measures in Panel A of Table 2.5 and the cross-sectional average bond market liquidity beta β^{CBLIQ} and transaction cost level c in Table A.2 of Appendix A.6, we can quantify the market liquidity risk and level premium as $\beta^{\text{CBLIQ}} \cdot \lambda_{\text{CBLIQ}}$ and $c \cdot \lambda_{\text{c}}$. The corporate bond market liquidity risk premium accounts for an expected excess return of about 0.91% (0.77%) p.a. and the level premium is responsible for an excess return of about 0.93% (0.83%) p.a. in case of the size-adapted Schultz (2001) (average bid-ask spread) measure.

Aside the corporate bond liquidity effects, we consistently find a significant premium for equity market risk and for equity market liquidity risk, confirming results of Bongaerts, de Jong, and Driessen (2017). However, the importance of equity market-specific effects compared to corporate bond liquidity effects varies strongly between settings employing a standard measure and those employing the size-adapted counterpart. While the sum of the equity effects of 1.20% and 1.15% dominate the sum of corporate bond effects of 0.57% and 0.61% for the standard Schultz (2001) and average bid-ask spread measure, respectively, this relation is reversed when moving from standard to size-adapted liquidity measures. Now, equity market-specific effects (0.99% to 1.00%) are much weaker than the corporate bond market liquidity risk and level premium (1.60% to 1.84%).[28] Interestingly, the reduction in equity premia is mainly driven by the equity liquidity risk premium which drops roughly by 30% to 50% when corporate bond liquidity is measured with the size-adapted liquidity measures. Therefore, our new approach shows that the corporate bond

[28]These premia are based on the non-overlapping triple sort. In the other three settings, bond market liquidity effects are also stronger than equity effects when liquidity is measured with our new approach.

market-specific effects are responsible for the major share of premia in expected bond excess returns.

2.5 Robustness

In this section, we perform two robustness checks. We first compare the explanatory power of the alternative approaches to account for the larger transaction costs of small trades with our size-adapted measures in the yield-spread setting of Section 2.3.1. Second, we show that our asset pricing results are robust when using a parametric functional form instead of the nonparametric version employed in the main analyses.

2.5.1 Simple Approaches to Account for Size-Dependence

A common approach in the literature to account for the volume-dependence of transaction costs is to discard all trades below a threshold, which is in most cases \$100,000. In a similar spirit, some researches calculate volume-weighted averages so that liquidity measures effectively resemble the transaction costs of large trades when they are available. We compare these two approaches to our size-invariant measurement approach.

To this end, we employ the yield-spread regressions from Section 2.3.1. Because yield spreads are already calculated on a volume-weighted basis, we can evaluate the increase in explanatory power for the size-weighted transaction cost measures immediately. Given the relative improvement compared to the baseline regression of 7.7% and 19.3% for the size-adapted measures in Table 2.3, volume-weighted transaction cost measures only lead to a relative improvement of 4.0% and 8.8% (results not tabulated to conserve space). Thus, the relative improvement of the simple, alternative approach is not only worse compared to our new approach but is also lower than when employing the unadjusted standard measures.

For the approach to discard trades below \$100,000, we estimate Equation (2.6) for a sample in which we calculate yield spreads and all control variables only from trades with volumes of at least \$100,000. Table 2.6 shows that the R^2_{adj} using only trades of at least \$100,000 increases by 7.5% and 25.2% for the Schultz (2001) and average bid-ask spread measure, respectively. To compare the standard approach based on all trades to the approach of deleting small trades, we include a specification in which we calculate transaction costs from all trades. In this setting, the relative improvements range between

Table 2.6: **Yield spread regressions: Trades with volumes $\geq \$100,000$**

This table reports results for the panel regression model (2.6) explaining monthly yield spread changes based on trades with volumes $\geq \$100,000$ with monthly changes in standard or size-adapted transaction cost measures. The control variables are trade size, number of trades, and amount outstanding. We further employ rating dummies as a proxy for credit risk. Specifications (1) and (5) are the baseline regressions that include only lagged yield spread changes and the control variables. In (2) and (6), we add standard transaction cost measures based on trades with volumes $\geq \$100,000$, while in (3) and (7) we add the measures based on all trade sizes. Finally, (4) and (8) incorporate the size-adapted measures, which, by construction, employ all trade sizes. We winsorize yield spread and transaction cost changes at the 1% and 99% level. Standard errors are clustered monthly and by bond. We test for differences between out-of-sample mean squared errors (MSE) using the test statistic proposed by Harvey, Leybourne, and Newbold (1997). Comparing R^2_{adj} or MSE, we first compare a specification to the baseline and second to its preceding specification. The t-statistics are given in parentheses. ** and * represent statistical significance at the 1% and 5% level.

	Schultz (2001)				Average bid-ask spread			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	0.0083	0.0084	0.0088	0.0090	0.0192	0.0198	0.0208	0.0211
	(0.32)	(0.32)	(0.34)	(0.35)	(0.51)	(0.54)	(0.58)	(0.60)
Δ(Yield spread)$_{i,t-1}$	0.1156	0.1152	0.1156	0.1158	0.1562*	0.1562*	0.1543*	0.1553*
	(1.82)	(1.82)	(1.82)	(1.83)	(2.16)	(2.18)	(2.19)	(2.26)
Δ(Standard measure)$_{i,t}$		6.3869**	8.3854**			22.1686**	23.1594**	
		(5.61)	(6.75)			(5.67)	(6.11)	
Δ(Adapted measure)$_{i,t}$				0.1037**				0.2175**
				(5.46)				(5.60)
Δ(Trade size)$_{i,t}$	-0.0110**	-0.0056	-0.0036	-0.0133**	-0.0186**	0.0014	-0.0011	-0.0238**
	(-3.18)	(-1.39)	(-0.87)	(-4.30)	(-3.93)	(0.20)	(-0.17)	(-5.51)
Δ(Trades)$_{i,t}$	0.0227**	0.0206**	0.0236**	0.0217**	0.0350**	0.0300**	0.0348**	0.0318**
	(5.55)	(5.41)	(5.70)	(5.53)	(5.13)	(4.94)	(5.34)	(5.12)
Δ(Amount outstanding)$_{i,t}$	-0.0366	-0.0342	-0.0213	-0.0224	-0.2642	-0.2549	-0.2588	-0.2464
	(-0.49)	(-0.45)	(-0.28)	(-0.30)	(-1.50)	(-1.43)	(-1.48)	(-1.44)
Δ(Rating dummies)$_{i,t}$	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2_{adj}	0.0571	0.0614	0.0647	0.0693	0.0536	0.0671	0.0714	0.0833
$\Delta(R^2_{adj})$		7.5%	13.3%/5.4%	21.4%/7.1%		25.2%	33.2%/6.4%	55.4%/16.7%
MSE	0.682	0.679	0.676	0.673	1.403	1.380	1.372	1.350
Δ(MSE)		-0.4%**	-0.9%**/-0.4%**	-1.3%**/-0.4%**		-1.6%**	-2.2%**/-0.6%**	-3.8%**/-1.6%**
		(5.51)	(7.42)/(3.58)	(7.98)/(4.20)		(9.92)	(11.56)/(3.11)	(12.73)/(6.70)
Observations			271,235				255,294	

13.3% and 33.2%. Once again, the size-adapted versions lead to the highest improvements of 21.4% and 55.4%.

Summarizing, both alternative approaches fail to better capture liquidity induced shifts in a bond's yield spread compared to our size-invariant measurement. Interestingly, both approaches even lead to worse explanatory powers than the standard equally-weighted transaction cost measures. These results show that the information in small trades is valuable and can be best exploited with our new approach.

2.5.2 Parametric Estimation of the Market-Wide Cost Function

A precise estimation of the market-wide transaction cost function is crucial for our approach. Besides the presented nonparametric estimation of Section 2.2.1, it is also possible to estimate the cost function with a parametric functional form. A parametric approach is easier to implement and faster to compute. Thus, we reexamine our findings on the corporate bond liquidity effects of Section 2.4 with size-adapted liquidity measures based on a parametric functional form.

Following Edwards, Harris, and Piwowar (2007), we employ the following parametric functional form to calculate transaction costs $c(vol)$ for a trade with volume vol:

$$c(vol) = c_0 + \frac{c_1}{vol} + c_2 \cdot \ln(vol) + c_3 \cdot vol + c_4 \cdot vol^2 + c_5 \cdot vol^3, \tag{2.10}$$

where $c_i \in \mathbb{R}$ for $i = 0, \ldots, 5$. Using this parametric function, we run the iterative two-stage weighted regressions described in Sections 2.2.2 and 2.2.3. The resulting transaction cost functions are depicted in Figure 2.5. Consistent with the nonparametric results of Figure 2.2, we find a monotonically decreasing relation for trade sizes up to about \$3 million and an increase for very large trades. However, the parametric forms suffer from a slightly oscillating pattern above \$1 million.

We employ the size-adapted measures based on the parametric cost function for our asset pricing tests. Across all four different settings in Table 2.7, we find a significantly positive corporate bond market liquidity risk and liquidity level premium for both the Schultz (2001) and the average bid-ask spread measure. These results are confirmed when adapting the repeat-sales measure of Section 2.4.4 using a parametric cost function. Thus, the parametric size-adapted measures consistently imply that both corporate bond market liquidity risk and liquidity level are priced.

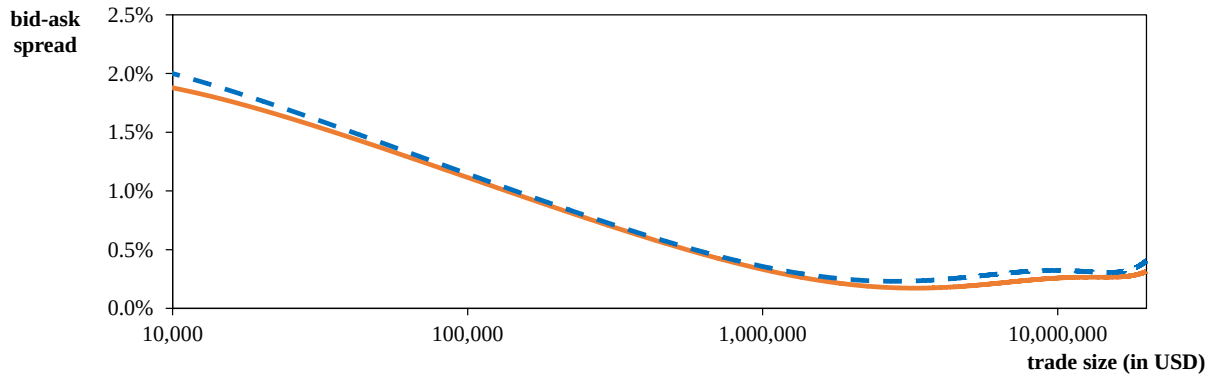


Figure 2.5: **Parametric transaction cost functions of U.S. corporate bonds**
The figure depicts parametric transaction cost functions for the size-adapted Schultz (2001) measure (orange solid line) and the size-adapted average bid-ask spread measure (blue dashed line). They are based on the parametric functional form

$$c(vol) = c_0 + \frac{c_1}{vol} + c_2 \cdot \ln(vol) + c_3 \cdot vol + c_4 \cdot vol^2 + c_5 \cdot vol^3.$$

The estimation procedure is described in Section 2.5.2.

2.6 Conclusion

In this chapter, we address the problem that common transaction cost measures for the bond market do not account for the fact that trading costs of small trades are typically much larger compared to large trades. We develop a simple two-stage liquidity measurement approach that accounts for the size dependence and aggregates trading costs from all trade sizes into one single value. Our approach is easily implementable, applicable to a broad variety of liquidity measures, and has no influence on a measure's data requirements. In this spirit, we adapt two standard measures of transaction costs, the Schultz (2001) and the average bid-ask spread measure, to our new approach.

Eliminating the dependence on changes in the average trade size leads to more precise measures of a bond's liquidity. Our new approach reduces the unsystematic noise in the liquidity estimates of scarcely traded bonds. We verify the lower estimation noise by examining the measures' capability to explain the embedded liquidity premium in corporate bond yield spreads. Lastly, in fire sale situations when changes in the average trade size become systematic, our size-adapted measures are able to detect the liquidity deterioration which standard measures fail to identify.

Finally, we examine the impact of our new liquidity measures on the asset pricing impli-

Table 2.7: **Fama-MacBeth regression: Parametric transaction cost function**
This table reports time series averages of the monthly results from the cross-sectional Fama-MacBeth regression (2.8). We use the size-adapted Schultz (2001) and average bid-ask measures based on the parametric functional form of Section 2.5.2 to calculate transaction costs. Additionally, we employ in the same way a size-adapted repeat-sales method. We present results for the overlapping and non-overlapping triple sort, the industry sort, and individual bonds (all settings are described in Section 2.4.3). The Fama-MacBeth t-statistics are calculated based on Newey and West (1987) standard errors with six lags and are given in parentheses. ** and * represent statistical significance at the 1% and 5% level.

	Non-overlapping triple sort			Overlapping triple sort		
	Schultz (2001)	Avg. bid-ask spread	Repeat-sales	Schultz (2001)	Avg. bid-ask spread	Repeat-sales
Intercept	-0.0003	0.0006	0.0011	-0.0017	0.0007	-0.0025
	(-0.20)	(0.26)	(0.43)	(-0.55)	(0.20)	(-0.70)
λ_{EQ}	0.0556**	0.0538**	0.0560**	0.0558**	0.0653**	0.0533**
	(5.34)	(4.61)	(5.59)	(3.34)	(4.60)	(3.29)
λ_{EQLIQ}	-0.0230**	-0.0275**	-0.0229**	-0.0438**	-0.0282*	-0.0320**
	(-3.29)	(-4.41)	(-3.89)	(-2.78)	(-2.53)	(-3.04)
$\lambda_{\text{CBLIQ}}^{\text{Size-adapted}}$	-0.0497**	-0.0434**	-0.0402**	-0.0510**	-0.0459*	-0.0517**
	(-4.20)	(-4.45)	(-3.67)	(-3.03)	(-2.36)	(-4.38)
$\lambda_{\text{c}}^{\text{Size-adapted}}$	0.0078**	0.0082**	0.0090**	0.0115**	0.0092*	0.0138**
	(3.89)	(4.92)	(4.00)	(2.78)	(2.45)	(3.68)
R^2_{adj}	70.2%	70.9%	72.2%	81.8%	81.7%	84.0%

	Industry sort			Individual bonds	
	Schultz (2001)	Avg. bid-ask spread	Repeat-sales	Schultz (2001)	Avg. bid-ask spread
Intercept	-0.0004	-0.0001	0.0059	0.0078**	0.0080**
	(-0.19)	(-0.07)	(1.58)	(5.89)	(6.18)
λ_{EQ}	0.0415**	0.0436**	0.0437**	0.0299**	0.0301**
	(5.58)	(5.85)	(5.27)	(3.80)	(4.16)
λ_{EQLIQ}	-0.0145**	-0.0133**	-0.0180**	-0.0150**	-0.0154**
	(-3.72)	(-3.04)	(-5.06)	(-3.02)	(-3.27)
$\lambda_{\text{CBLIQ}}^{\text{Size-adapted}}$	-0.0356**	-0.0422**	-0.0422**	-0.0234**	-0.0275**
	(-2.91)	(-3.23)	(-3.20)	(-2.79)	(-2.96)
$\lambda_{\text{c}}^{\text{Size-adapted}}$	0.0109**	0.0104**	0.0071*	0.0061**	0.0055**
	(4.80)	(4.33)	(2.53)	(6.58)	(7.56)
R^2_{adj}	72.7%	70.3%	78.1%	39.9%	37.8%

cations of corporate bond market liquidity risk and individual liquidity level for expected corporate bond excess returns. Using established standard measures of the literature, we show that they lead to inconsistent pricing patterns and that results depend critically on the portfolio selection and beta estimation approach. In contrast, when using the size-adapted liquidity measures, bonds with higher transaction costs or with a stronger sensitivity to corporate bond market liquidity consistently earn higher expected excess returns. This finding does not depend on the construction of test assets and the beta estimation procedure. Thus, our new measurement approach uncovers that U.S. corporate bonds pay a liquidity premium for both their individual liquidity and their exposure to market-wide bond liquidity risk. Given the economically large size of these premia, investors should consider both effects for their optimal portfolio choice.

Chapter 3

Expected Bond Liquidity

3.1 Introduction

A basic principle of financial economics is that expectations about future market conditions influence decisions. Because liquidity is an elusive concept and the literature has found multifaceted relations with other market factors, forming expectations about future liquidity is difficult. In the absence of a generally accepted forecasting approach, market participants are left alone to aggregate and extrapolate the available information when they assess future liquidity. For example, investors require the liquidity at the future time of sale to evaluate the expected payoff of a trading strategy. Regulators and central banks monitor the expected development of liquidity very closely to take timely countermeasures. Finally, issuing companies react to their bonds' expected liquidity deteriorations to avoid distress arising from worsening refinancing conditions (see, e.g., He and Xiong, 2012).

We are not aware of a (sophisticated) forecasting model for individual bond liquidity in the academic literature.[29] Instead, researchers employ the naïve assumption that a bond's liquidity today is the best estimator for its liquidity tomorrow. We fill this gap and introduce a forecasting model for individual bond liquidity. Our objective is to employ the information available to a contemporary forecaster to most precisely estimate a bond's liquidity in the month ahead. To this end, we exploit the large pool of drivers of liquidity from the literature and dynamically select each month the subset of predictors that offers the best predictive power given the current information set.

[29] Forecasting liquidity at the market level, Boyarchenko, Giannone, and Shachar (2019) find that autoregressive models are hard to beat.

Our forecasting procedure combines elements from machine learning with the transparency of a simple linear model. In each month, our algorithm performs the following steps based on a rolling window of the previous twelve months. First, we use three different approaches to select the predictor variables that have the strongest forecasting power for the most recent past. We use elastic net, a variant of stepwise regression, and a method that relies on significant relations within the calibration window. We then calibrate a simple linear model to the selected variables. To further increase the predictive accuracy, we combine the forecasts from the three selection approaches to an average forecast (see, e.g., Rapach, Strauss, and Zhou, 2010). We implement the prediction model on the U.S. corporate bond market for the simple average bid-ask spread measure of Hong and Warga (2000) using transaction data from Enhanced TRACE for the period from October 1, 2004 to June 30, 2017. Note that the procedure can be easily applied to any liquidity measure and we consider a more advanced liquidity measure in the robustness section.

We evaluate the performance of the new prediction model relative to the literature's naïve approach on the basis of a direct and an indirect comparison. First, we compare the forecasting errors in an out-of-sample setting for our forecasting model and the naïve prediction. We find that our new model outperforms the naïve prediction model in every month of our observation period from 2004 to 2017. Interestingly, the largest performance improvements occur during the financial crisis. Overall, our forward-looking approach reduces the average forecasting error by about 19%. Second, in the indirect comparison, we show that the predictions of our new model better explain the premium for expected liquidity in corporate bond yields. We exploit Amihud and Mendelson's (1986) finding that investors require higher expected returns for assets that trade at higher (future) transaction costs. Following their guidance, we regress, in a panel setting, monthly yield-spread changes on changes in expected liquidity and a set of control variables. We compare the results of this analysis using expected liquidity from our forecasting model with the results when using the naïve assumption that a bond's liquidity today is the best estimator for its liquidity tomorrow. We find a much higher sensitivity of yield spreads to expected liquidity combined with a higher explanatory power when we use our sophisticated model. The about seven times higher sensitivity of yield spreads to changes in expected liquidity indicates that a naïve approach strongly underestimates the influence of liquidity on financing costs.

Finally, we leverage the forward-looking nature of our approach and shed light on the strategic behavior of investors in corporate bond funds. When selling their shares, investors in mutual funds usually receive the net asset value as of the time of sale. Costly

portfolio readjustments, however, happen at a later date and lead to negative externalities for the investors who stay in the fund. The resulting first-mover advantage can lead to 'runs' on the fund similar to bank runs. Consequently, flows out of poorly performing funds are exacerbated when the funds' portfolio is illiquid (Goldstein, Jiang, and Ng, 2017). Given this background, we analyze whether investors anticipate liquidity deteriorations and incorporate this information into their redemption decisions. To discriminate between investors who react to observed liquidity and those that actively form expectations, we regress, in a panel setting, monthly corporate bond fund flows on a fund's current liquidity and its expected liquidity change. For funds with a negative performance, we find evidence consistent with investors indeed acting on expected liquidity deteriorations. This anticipation channel reinforces the established effect that investors oversell poorly performing funds with currently illiquid holdings. Intuitively, both effects become more pronounced if a fund's performance gets worse.

We do not claim that market participants have formed their expectations in exact accordance with our prediction model, and we concede that their prediction approaches may even work better. However, market participants' forecasts are likely correlated with our predictions. Indeed, our results indicate that the forecast of the marginal investor is correlated stronger with our prediction than with a bond's current liquidity as the naïve prediction. This finding contrasts the practice in the literature to employ a bond's current liquidity when, formally, expected liquidity is required. For example, in asset pricing applications on bond liquidity, essentially all papers use a bond's current liquidity (see, e.g., Bao, Pan, and Wang, 2011; Friewald, Jankowitsch, and Subrahmanyam, 2012; Dick-Nielsen, Feldhütter, and Lando, 2012; Bongaerts, de Jong, and Driessen, 2017).[30] Moreover, our robustness test using a different forecasting methodology based on a random forest model reveals that our findings do not depend on the exact procedure used to calculate forecasts.

The remainder of the chapter is structured as follows. In Section 3.2, we start with presenting the pool of our candidate predictors in Section 3.2.2 and then introduce our dynamic forecast approach in Section 3.2.3. We evaluate the forecast performance of our new approach in Sections 3.2.4 and 3.2.5. Section 3.3 examines strategic decisions on early redemptions of bond fund investors when a struggling fund's liquidity is expected to deteriorate. We conclude in Section 3.5.

[30]Some papers argue that (market) liquidity follows an AR-1 process (see, e.g., Amihud, 2002). If an AR-1 process is the best model to describe liquidity movements, today's liquidity contains all information and could be used as proxy for tomorrow's expected liquidity. Our results indicate that this is not the case.

3.2 Predicting Bond Liquidity

In this section we introduce our liquidity prediction approach. We identify a set of *candidate* predictor variables for which a close connection with bond liquidity has been documented in the literature. For each point in time, our forecasting algorithm selects from this set of candidates the variables that have the highest predictive power in the most recent past. Our goal is to find the best-performing model from the perspective of a contemporary observer, mitigating the impact of a look-ahead bias that would arise if the variables were selected based on full-sample information. We compare the performance of our prediction model with a naïve benchmark model assuming a bond's liquidity in the next month is unchanged from today. Such a naïve forecast is exactly what researchers implicitly do when they use the currently prevailing liquidity in their applications instead of the expected liquidity actually required. Hence, we also examine whether our measure for predicted liquidity is able to better explain changes in bond yield spreads compared to the naïve approach.

3.2.1 Data and Liquidity Measure

Our analysis is based on bond transaction data from Enhanced TRACE from October 1, 2004 to June 30, 2017 (see Appendix B.1 for details). Bond characteristics, rating histories, and outstanding amounts are from Reuters Eikon and Bloomberg. We implement our liquidity forecast for the commonly used average bid-ask spread measure of Hong and Warga (2000) and for the more advanced liquidity measure of Chapter 2 that incorporates the dependence of transaction costs on trade size in the robustness section. The average bid-ask spread for bond i in month t can be calculated as

$$AvgBidAsk_{i,t} = Avg\left[\frac{\overline{P_{i,d}^{buy}} - \overline{P_{i,d}^{sell}}}{0.5 \cdot \left(\overline{P_{i,d}^{buy}} + \overline{P_{i,d}^{sell}}\right)}\right], \tag{3.1}$$

where $\overline{P_{i,d}^{buy/sell}}$ is the average of all buy or sell trades in bond i on day d.

3.2.2 Drivers of Liquidity

Despite bond liquidity being persistent (see, e.g., Chordia, Sarkar, and Subrahmanyam, 2005; Acharya, Amihud, and Bharath, 2013), a large body of empirical literature shows that it varies predictably over the lifetime of a bond. Early studies find that bonds are typically most liquid directly after issuance and get more illiquid when they age (see, e.g., Warga, 1992; Hong and Warga, 2000). Bonds with a higher outstanding amount and bonds that trade more frequently have lower transaction costs (see, e.g., Edwards, Harris, and Piwowar, 2007; Bao, Pan, and Wang, 2011; Jankowitsch, Nashikkar, and Subrahmanyam, 2011). Riskier bonds with a higher duration and more credit risk are typically less liquid than comparable bonds with lower risks (see, e.g., Mahanti, Nashikkar, Subrahmanyam, Chacko, and Mallik, 2008; Hotchkiss and Jostova, 2017). Chordia, Roll, and Subrahmanyam (2000) show for the stock market that individual trading costs move together with market and sector specific trading costs. Liquidity is also related to broader measures of market functioning. For example, Chordia, Sarkar, and Subrahmanyam (2005) and Goyenko and Ukhov (2009) identify bond market performance, volatility, order imbalance, and spillover effects from the stock market as driving factors of bond market liquidity. Additionally, these authors find that macroeconomic variables such as monetary policy, inflation, or industrial production have a significant connection to liquidity.[31]

Based on this literature, we build our set of candidate predictors to forecast next period's bid-ask spreads $\widehat{AvgBidAsk}_{i,t+1}$. Given the high persistence of individual and market liquidity, we naturally include a bond's bid-ask spread in the current month t. Because many bonds trade very infrequently leading to potentially noisy liquidity measures, we additionally include the moving average of the liquidity measure from the previous twelve months. As liquidity and its standard deviation are closely related (Dick-Nielsen, Feldhütter, and Lando, 2012), we also consider the standard deviation of the daily bid-ask spread measure. Next, we include a bond's age, its duration, and its (log-transformed) outstanding amount.[32] We capture trading activity with the logarithms of the average trade size and the total trading volume. Following Chordia, Sarkar, and Subrahmanyam (2005), we incorporate a bond's monthly return and order imbalance as possible predic-

[31] We do not consider bond characteristics that usually do not change during a bond's life such as the coupon, the original time-to-maturity, embedded options, or industry effects (see, e.g., Edwards, Harris, and Piwowar, 2007; Hotchkiss and Jostova, 2017; Jankowitsch, Nashikkar, and Subrahmanyam, 2011; Mahanti, Nashikkar, Subrahmanyam, Chacko, and Mallik, 2008). We capture the effects of such time-invariant variables through a bond's lagged liquidity.

[32] Note that our data for amount outstanding from Reuters Eikon includes reopenings, repurchases, and other (early) redemptions.

tors. We measure bond order imbalance as the difference between a bond's buying and selling dollar volume normalized with total trading volume. Regarding credit risk, we use the average numerical bond rating of the three rating agencies S&P, Moody's, and Fitch[33] and the five-year CDS spread from Markit.

Given the strong commonality of individual liquidity with market-wide liquidity, we include aggregate corporate bond market liquidity. We measure monthly market liquidity as the equally-weighted average bid-ask spread across all bonds in the sample. Before aggregating, we winsorize spreads at the 1% and 99% levels. In the spirit of Chordia, Roll, and Subrahmanyam (2000), we include a more granular aggregate liquidity measure, which is motivated by the fact that bonds with similar characteristics might be driven by the same market forces as they are to some extent substitutes to each others. To this end, we perform an independent triple sort in each month. Each portfolio represents a bond segment and we use the average portfolio bid-ask spread for each bond in the portfolio as candidate predictor. We follow Bongaerts, de Jong, and Driessen (2017) and Downing, Underwood, and Xing (2005) by sorting on a bond's average rating (quartiles), on its amount outstanding (terciles), and on its time to maturity (terciles), leading to 36 different segments.[34]

Following Goyenko and Ukhov (2009), we include short- and long-term market returns using the one-month and twelve-month return of the Barclay's U.S. corporate bond index. We also employ market volatility and order imbalance (see Chordia, Sarkar, and Subrahmanyam, 2005). We capture bond market volatility via CBOE's 10-year U.S. treasury note volatility index (TYVIX) and via the annualized realized volatility of Barclay's U.S. corporate bond index within month t. Market order imbalance is calculated as the difference between monthly aggregate buying and selling dollar volume normalized with total trading volume. Considering spillover effects from the equity market, we use the one-month and twelve-month return of the S&P 500. We also include equity market volatility via CBOE's volatility index (VIX) as well as equity market liquidity.[35] Following common practice, we use the Amihud (2002) price impact measure to approximate equity market liquidity.[36]

[33]We transform the ratings to integer numbers (AAA: 1, ... D: 22).

[34]As a robustness check, we also test a dependent version of the triple sort in which we first sort on rating, then on amount outstanding, and last on time to maturity. Additionally, we test an alternative ordering following Downing, Underwood, and Xing (2005) and sort on time to maturity, rating, and finally on the amount outstanding. In both settings, results remain qualitatively the same.

[35]We do not include stock market order imbalance due to data limitations.

[36]We calculate a monthly measure as the equally-weighted mean from all stocks with share codes of 10 and 11 in CRSP. We exclude observations on days without trading and require at least three days with positive trading volume per month in a stock. Further, we exclude shares traded at NASDAQ and winsorize the monthly cross-section of individual Amihud (2002) measures at the 5% and 95% level.

Regarding broader (macro) economic factors, we include the inflation rate using data from the OECD as well as the one-month and six-month TED spread, the federal funds rate to approximate monetary policy, and yearly industrial production growth using data from the Federal Reserve Bank of St. Louis.[37]

We report descriptive statistics for our set of candidate predictor variables in Table 3.1. Panel A shows average cross-sectional statistics for the variables with both time-series and cross-sectional variation. The average bond has a bid-ask spread of 1.3%, a monthly return of 0.5%, an age of roughly four years, and a duration of roughly six years. Regarding trading activity, the average bond has a log-transformed total trading volume of 17.6 (corresponding to about $45 million) and exhibits a slightly negative order imbalance. Regarding credit quality, the average bond has a rating of about 8 (corresponding to BBB+) and a credit default swap spread of 1.7% p.a. Overall, all variables show a strong variation in the cross-section. Panel B reports time series statistics for our market and macroeconomic variables. For both bond and equity market liquidity, the 95%-percentile shows that during crises, market-wide trading costs and price impact have been about two times as large as on average. Corporate bond and equity markets generate an average yearly return of about 5% and 9% during our sample period. In aggregate, the bond market exhibits a slightly negative order imbalance. Regarding macroeconomics, the average inflation rate is about 2% and industrial production growths by roughly 0.7% per year. Again, all variables show a strong variation.

Table 3.2 presents correlations for our set of candidate predictors. We additionally include next month's liquidity to get a first insight on which variables might have predictive power. As expected, we observe a strong correlation of 0.71 between current liquidity and next month's liquidity. The high persistence is also confirmed by the correlation of 0.73 between next month's liquidity and the moving average of the previous twelve months. This finding shows that a naïve forecast of future liquidity using current liquidity is not a bad starting point. However, there are several other variables that also show a strong correlation with next month's liquidity. Especially the trading activity variables exhibit a promising relation. Consistent with the literature, we find that a higher average trade size and a higher total trading volume are associated with lower bid-ask spreads. Also as expected, we find a negative correlation of -0.43 between outstanding amount and next month's liquidity and confirm the well-known negative relation between credit quality and future liquidity (see He and Xiong, 2012). Consistent with our ex-ante expectation that

[37]Note that data on inflation and industrial production becomes available with a time lag of one month. Thus, we lag these two variables by one month.

Table 3.1: **Descriptive statistics for candidate predictors**
This table shows descriptive statistics for the set of candidate predictors (see Section 3.2.2). In Panel A, we report average cross-sectional statistics for transaction cost variables, bond characteristics, trading activity, and credit quality variables that have both time-series and cross-sectional variation. To this end, we first calculate each month the cross-sectional statistics and then average over time. In Panel B, we report time series statistics for the predictor variables on the bond market, the equity market, and the broader economic environment. We measure bond liquidity using the average bid-ask spread measure.

Panel A: Variables with time-series and cross-sectional variation

	Mean	Std. dev.	$Q_{5\%}$	$Q_{25\%}$	$Q_{50\%}$	$Q_{75\%}$	$Q_{95\%}$
Current liquidity (%)	1.30	1.22	0.05	0.42	0.96	1.87	3.65
12-month mov. avg. Liquidity (%)	1.32	1.02	0.19	0.54	1.06	1.89	3.29
Segment liquidity (%)	1.29	0.67	0.44	0.81	1.13	1.67	2.65
Volatility of liquidity (%)	0.49	0.61	0.00	0.01	0.32	0.72	1.67
Bond return (%)	0.48	2.88	-3.35	-0.59	0.39	1.49	4.52
Age (years)	4.30	3.71	0.43	1.73	3.43	5.77	11.81
Duration (years)	6.07	4.64	0.57	2.47	4.87	8.34	15.51
Amount outstanding (log USD)	19.19	1.65	15.84	18.23	19.64	20.35	21.23
Average trade size (log USD)	12.36	1.51	9.65	11.20	12.74	13.53	14.27
Total trading volume (log USD)	17.63	2.26	13.41	16.10	18.16	19.31	20.56
Bond order imbalance (%)	-0.61	35.05	-65.43	-16.95	-0.41	16.14	62.43
Rating (1: AAA,..., 22: D)	7.75	3.40	2.42	5.52	7.38	9.74	14.19
CDS spread (%)	1.73	3.14	0.28	0.47	0.84	1.68	6.12

Panel B: Variables with time-series variation

	Mean	Std. dev.	$Q_{5\%}$	$Q_{25\%}$	$Q_{50\%}$	$Q_{75\%}$	$Q_{95\%}$
Bond market liquidity (%)	1.29	0.48	0.73	0.86	1.29	1.55	2.24
1-month bond market return (%)	0.43	1.62	-1.84	-0.45	0.45	1.37	2.77
12-month bond market return (%)	5.44	6.43	-4.26	1.65	5.14	8.15	18.68
Bond market order imbalance (%)	-0.64	2.29	-4.28	-1.90	-0.60	0.86	2.54
Bond market volatility (%)	4.79	1.80	2.75	3.52	4.34	5.72	8.31
TYVIX (%)	6.30	1.93	4.30	4.92	5.66	7.03	10.49
Stock market liquidity	0.042	0.018	0.022	0.028	0.040	0.050	0.074
1-month stock market return (%)	0.76	4.02	-7.03	-1.49	1.27	3.27	6.78
12-month stock market return (%)	9.38	15.99	-32.57	4.91	12.11	17.37	30.20
VIX (%)	19.00	8.98	11.26	13.43	16.19	21.61	36.53
Inflation (%)	2.06	1.47	-0.20	1.15	1.99	3.17	4.31
1-month TED spread (%)	0.38	0.46	0.12	0.16	0.20	0.36	1.40
6-month TED spread (%)	0.52	0.39	0.23	0.28	0.35	0.60	1.31
Federal funds rate (%)	1.36	1.86	0.08	0.12	0.19	2.39	5.25
Industrial production growth (%)	0.69	4.70	-11.63	-0.84	2.26	3.25	5.31

Table 3.2: **Liquidity and candidate predictor correlation matrix**
This table reports correlations for next month's bond liquidity and the candidate predictors (see Section 3.2.2) based on panel data. We measure bond liquidity using the average bid-ask spread measure.

	Next month's liquidity	Current liquidity	12-month liquidity	Segment liquidity	Volatility of liquidity	Bond return	Age	Duration	Amount outstanding	Average trade size	Total trading volume	Bond order imbalance	Rating	CDS spread	Bond market liquidity
Next month's liquidity	1	0.71	0.73	0.63	0.40	0.00	0.17	0.32	-0.43	-0.48	-0.31	-0.01	0.09	0.27	0.36
Current liquidity		1	0.75	0.65	0.47	0.02	0.17	0.31	-0.43	-0.50	-0.32	-0.03	0.09	0.28	0.37
12-month liquidity			1	0.69	0.39	0.07	0.24	0.34	-0.54	-0.60	-0.40	0.00	0.10	0.28	0.37
Segment liquidity				1	0.32	0.05	0.10	0.41	-0.62	-0.55	-0.50	-0.02	0.11	0.23	0.56
Volatility of liquidity					1	0.03	0.08	0.20	-0.04	-0.11	0.04	-0.01	0.19	0.28	0.28
Bond return						1	0.02	0.04	-0.02	0.00	0.00	-0.05	0.04	0.00	0.05
Age							1	-0.02	-0.12	-0.13	0.00	0.01	0.09	0.09	0.00
Duration								1	-0.08	-0.06	-0.03	-0.01	-0.07	-0.04	-0.02
Amount outstanding									1	0.81	0.89	0.00	0.00	-0.12	-0.24
Average trade size										1	0.80	0.01	0.15	0.01	-0.22
Total trading volume											1	0.01	0.10	0.06	-0.17
Bond order imbalance												1	0.01	0.00	-0.01
Rating													1	0.50	-0.05
CDS spread														1	0.20
Bond market liquidity															1
Bond market return (1m)															
Bond market return (12m)															
Bond market order imbalance															
Bond market volatility															
TYVIX															
Stock market liquidity															
Stock market return (1m)															
Stock market return (12m)															
VIX															
Inflation															
TED spread (1m)															
TED spread (6m)															
Federal funds rate															
Industrial production growth															

Table 3.2 continued

	Bond market return (1m)	Bond market return (12m)	Bond market order imbalance	Bond market volatility	TYVIX	Stock market liquidity	Stock market return (1m)	Stock market return (12m)	VIX	Inflation	TED spread (1m)	TED spread (6m)	Federal funds rate	Industrial production growth
Next month's liquidity	0.01	-0.02	-0.05	0.19	0.25	0.33	-0.08	-0.20	0.24	0.11	0.21	0.21	0.14	-0.14
Current liquidity	0.03	-0.01	-0.06	0.19	0.26	0.34	-0.06	-0.21	0.25	0.09	0.20	0.21	0.13	-0.15
12-month liquidity	0.07	0.15	-0.03	0.15	0.22	0.33	0.00	-0.12	0.19	0.04	0.10	0.12	0.12	-0.15
Segment liquidity	0.05	-0.02	-0.08	0.30	0.40	0.52	-0.10	-0.32	0.40	0.13	0.30	0.33	0.18	-0.25
Volatility of liquidity	0.02	-0.04	-0.03	0.17	0.22	0.27	-0.05	-0.20	0.22	0.04	0.15	0.18	0.06	-0.16
Bond return	0.37	0.10	-0.06	-0.07	0.02	0.08	0.18	-0.04	0.03	-0.12	-0.11	-0.01	-0.04	-0.09
Age	0.01	0.03	0.00	0.03	0.03	-0.01	0.00	0.00	0.04	-0.03	-0.01	0.03	-0.06	-0.02
Duration	0.01	0.02	0.00	0.01	0.01	-0.02	0.00	0.00	0.01	-0.03	-0.01	0.01	-0.04	-0.01
Amount outstanding	-0.01	-0.06	0.04	-0.02	-0.06	-0.21	0.01	0.02	-0.04	-0.15	-0.10	0.00	-0.23	-0.03
Average trade size	-0.01	-0.05	0.04	-0.02	-0.06	-0.20	0.01	0.02	-0.04	-0.15	-0.10	-0.01	-0.22	-0.04
Total trading volume	0.00	-0.03	0.03	0.01	-0.01	-0.16	0.01	0.02	0.01	-0.13	-0.08	0.01	-0.23	-0.03
Bond order imbalance	-0.02	0.01	0.05	-0.01	-0.01	-0.02	-0.01	0.01	-0.01	0.00	-0.01	-0.02	0.00	0.01
Rating	0.00	0.00	0.02	-0.05	-0.06	-0.05	0.02	0.03	-0.06	-0.02	-0.04	-0.05	0.00	0.00
CDS spread	0.03	-0.06	-0.03	0.16	0.20	0.21	-0.05	-0.19	0.22	0.00	0.12	0.21	-0.04	-0.15
Bond market liquidity	0.10	-0.02	-0.15	0.53	0.72	0.93	-0.17	-0.57	0.70	0.22	0.52	0.58	0.33	-0.43
Bond market return (1m)	1	0.20	-0.10	-0.16	0.05	0.15	0.28	-0.09	0.10	-0.20	-0.24	-0.03	-0.09	-0.14
Bond market return (12m)		1	0.07	-0.14	-0.13	-0.09	0.20	0.47	-0.14	-0.16	-0.32	-0.32	-0.20	0.17
Bond market order imbalance			1	-0.14	-0.14	-0.13	0.01	0.10	-0.13	-0.08	-0.14	-0.12	-0.07	-0.03
Bond market volatility				1	0.83	0.48	-0.26	-0.52	0.71	-0.11	0.38	0.57	-0.21	-0.40
TYVIX					1	0.66	-0.23	-0.62	0.84	-0.07	0.42	0.67	-0.21	-0.50
Stock market liquidity						1	-0.06	-0.64	0.69	0.14	0.37	0.52	0.25	-0.45
Stock market return (1m)							1	0.23	-0.33	-0.22	-0.31	-0.25	-0.06	0.02
Stock market return (12m)								1	-0.63	0.12	-0.38	-0.68	-0.03	0.64
VIX									1	-0.03	0.42	0.70	-0.22	-0.47
Inflation										1	0.46	0.14	0.52	0.48
TED spread (1m)											1	0.75	0.41	-0.12
TED spread (6m)												1	0.00	-0.48
Federal funds rate													1	0.19
Industrial production growth														1

older bonds are associated with lower liquidity, we find a positive correlation of 0.17 between age and next month's liquidity. In the same spirit, a positive correlation of 0.32 indicates that a higher duration leads to higher bid-ask spreads. Regarding market and macroeconomic variables, we find aggregate bond market and stock market liquidity to have the highest correlation with next month's liquidity of 0.36 and 0.33, respectively, confirming the co-movement between individual and market liquidity. Not surprisingly, our more granular aggregate liquidity measure (segment liquidity) with a correlation of 0.63 is even more strongly connected to a bond's liquidity in the next month. Interestingly, the sign and the magnitude of all correlations remain comparable when we consider current liquidity instead of next month's liquidity.

3.2.3 Prediction Model

Starting with the set of candidate predictors, we develop our forecast methodology. We implement an estimation procedure that exploits information up to time t to forecast liquidity in month $t+1$. In each month, we include only those variables in the model that increase the predictive power for the most recent past so that a contemporary forecaster would have been able to use the same information. The forecast is then based on the linear model

$$\widehat{AvgBidAsk}_{i,t+1} = \widehat{\alpha}_t + \sum_{m \in M_t} \widehat{\beta}_{m,t} \cdot \text{predictor}_{m,i,t}, \tag{3.2}$$

where, for each month t, we determine the set of predictor variables M_t and parameters $\widehat{\alpha}_t$ and $\widehat{\beta}_{m,t}$ based on a rolling window of twelve months $t-12, ..., t$ and a two-step procedure. In the first step, we select those variables M_t that have the highest predictive power within the twelve-month window. Therefore, M_t adapts to new information as it becomes available (see, e.g., Chinco, Clark-Joseph, and Ye, 2019; Pesaran and Timmermann, 1995, for a similar argument) and naturally accommodates structural changes in the relation between variables.[38] In the second step, we calibrate $\widehat{\alpha}_t$ and $\widehat{\beta}_{m,t}$ for the selected predictors. To mitigate the impact of outliers on variable selection and calibration, we winsorize the data at the 1% and 99% levels on a monthly basis.

We employ three different methods to determine the predictor set M_t in the first step.

[38]There are several alternatives to our approach when implementing a dynamic estimation procedure. First, one can use a recursive scheme instead of a rolling window. Second, the literature uses various lengths for rolling windows from twelve months up to five years (see, e.g, Fama and MacBeth, 1973; Kacperczyk, Nieuwerburgh, and Veldkamp, 2014). We find in unreported results that the higher flexibility of a rolling scheme and a window length of twelve months is more capable to adjust for structural changes and outperforms alternative specifications.

Two of the three selection methods are based on out-of-sample cross-validation. For these, we employ a holdout procedure in which the data of the first eleven months is used to train the model and the last month is used to validate the model's prediction performance. The predictor variables that lead to the lowest error on the validation set are then selected for M_t. Our first selection algorithm is a variant of stepwise regression (see, e.g., Agarwal and Naik, 2004; Titman and Tiu, 2011). Here, the algorithm adds variables that decrease the mean squared error (MSE) on the validation set until no other variable leads to further improvement. After each addition, the algorithm checks if dropping one of the selected variables decreases the model's error. The second selection algorithm is the elastic net procedure, which uses a combined penalty function of LASSO and ridge methods (see, e.g., Kozak, Nagel, and Santosh, 2020; Panopoulou and Vrontos, 2015). Again, the parameters controlling the number of variables included in the model are set to minimize the MSE on the validation data set. The third selection method relies on detecting stable predictive relations in-sample (see, e.g., Chernobai, Jorion, and Yu, 2011). The model is estimated using all candidate predictors and M_t simply contains those variables for which we observe a predictive relation with a significance level of lower than 5%, where we cluster standard errors by bond.[39] In the second step, we then use the full twelve months of data to calibrate the model on the predictors selected by the three selection methods.

For the three different methods, we can now calculate bond i's expected liquidity in the next month $\widehat{AvgBidAsk}_{i,t+1}$ using Equation (3.2) and the predictors' values in month t.[40] Rapach, Strauss, and Zhou (2010) find that combining predictions of individual models often results in superior performance. We follow these authors and average the three forecasts to arrive at our final prediction of next month's liquidity.[41]

In the robustness section 3.4.1, we also test an alternative prediction approach based on a random forest model. Albeit the machine learning model brings a slight improvement in terms of forecast accuracy, it should be noted that the relation of predictors to next month's liquidity and their economic significance is notoriously difficult to interpret within such approaches.

[39] Table 3.2 shows that many possible predictors are strongly connected. To mitigate multicollinearity issues, we exclude all variables with a variance inflation factor of 10 or higher (see, e.g., Liu and Ritter, 2011).

[40] In the rare case that $\widehat{AvgBidAsk}_{i,t+1}$ is negative, we set it to 0.

[41] The econometric literature finds that the simple average of different forecasts often outperforms more sophisticated weighting schemes (the phenomenon is called 'Forecast Combination Puzzle', see, e.g, Smith and Wallis, 2009). Combining the information of the three models generates the best predictions for our sample. However, the individual models also generate good results.

3.2.4 Prediction Results

We compare the accuracy of our new forecast methodology to the naïve benchmark model $\widehat{AvgBidAsk}_{i,t+1}^{\text{naïve}} = AvgBidAsk_{i,t}$. As discussed before, researchers that apply today's liquidity for applications that formally require future expected liquidity implicitly employ such a naïve model. The forecast evaluation criterion is the root mean mean squared error (RMMSE)

$$RMMSE^{\text{naïve/fc}} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_t}\sum_{i=1}^{n_t}\left(AvgBidAsk_{i,t+1} - \widehat{AvgBidAsk}_{i,t+1}\right)^2},$$

where $AvgBidAsk_{i,t+1}$ is the realized liquidity of bond i in month $t+1$ and $\widehat{AvgBidAsk}_{i,t+1}$ is the prediction of the naïve benchmark or the forecast model for month $t+1$ based on data from t. n_t is the number of bonds in month t for which we can assess liquidity in month $t+1$. Finally, T is the number of months in our observation period. As we employ a rolling window of twelve months to calibrate our model, predictions start in November 2005 and end in June 2017.

The average forecasting errors of the naïve benchmark and our forecast model are reported in Panel A of Table 3.3. Using a bond's current liquidity as the naïve forecast leads on average to an error of 91 basis points. In comparison, our linear combination model only has a forecasting error of 74 basis points, which corresponds to a relative outperformance of roughly 19%.[42] To get a better understanding of the superior performance of our forecasting model, we plot the time series of the monthly root mean squared error (RMSE) in Panel A of Figure 3.1. The performance of the new model surpasses the benchmark in each month of our observation period. Interestingly, while the highest prediction errors occur during the financial crisis, the largest improvement of the predictive accuracy also seems to coincide with the crisis period. Because this period is also the period with the highest average bid-ask spread, we plot the improvement in the RMSE of our forecasting model relative to the average bid-ask spread in Panel B of Figure 3.1. This figure shows that the outperformance compared to the naïive benchmark is relatively stable over time. Remarkably, the two months with the highest relative outperformance are within the financial crisis (March and December 2008). Thus, the application of our new model is especially advisable during times of liquidity stress.

[42] In Chapter 4, we introduce a testing procedure to compare the mean prediction error of forecast models in a panel setting. Applying this test yields that the linear combination model's stronger performance is significant at the 5% level.

Table 3.3: **Forecasting model**
Panel A of this table reports root mean mean squared errors (RMMSE) for the forecasting model and the naïve benchmark. The forecasting model is described in Section 3.2.3. The naïve forecast for a bond's liquidity in the next month equals the realization of today. We measure bond liquidity using the average bid-ask spread measure. The full model includes all candidate predictors from Section 3.2.2, whereas the restricted model only includes variables that are directly related to liquidity (see Section 3.2.4). Panel B shows statistics for the variables that are selected in the forecasting approach for the full model. Variables are ranked by the percentage of months for which they are included. Additionally, we report the percentage of months the variables have a positive and negative coefficient $\widehat{\beta}$ given that they are included and their economic significance (in bps). The economic significance is calculated as the average of the monthly product of the parameter estimate and the variable's standard deviation within the 12-month rolling window used for the model calibration. For the calculation of economic significance, variables that are not selected have a coefficient of 0. All statistics are based on the average parameter across the three individual selection approaches that form the combination model.

Panel A: Forecast performance

	Full model		Restricted model	
	RMMSE	Δ to naïve	RMMSE	Δ to naïve
Forecasting model	0.74	-18.71%	0.80	-18.81%
Naïve benchmark model	0.91		0.99	
Observations	230,790		511,465	

Panel B: Selected variables

Rank	Effect	% included	% positive	% negative	Econ. significance
1	12-month liquidity	99.8%	100.0%	0.0%	41.8
2	Current liquidity	99.8%	100.0%	0.0%	30.4
3	Duration	92.9%	100.0%	0.0%	6.7
4	Average trade size	89.8%	0.0%	100.0%	-5.3
5	CDS spread	80.7%	99.3%	0.7%	3.9
6	Segment liquidity	76.7%	98.5%	1.5%	5.1
7	Total trading volume	72.4%	100.0%	0.0%	4.5
8	Volatility of liquidity	71.7%	100.0%	0.0%	2.7
9	Amount outstanding	65.0%	1.4%	98.6%	-3.4
10	Bond return	57.6%	14.2%	85.8%	-1.5
11	Stock market return (1m)	45.7%	17.1%	82.9%	-1.4
12	Age	42.9%	43.5%	56.5%	-0.2
13	Rating	36.0%	31.1%	68.9%	-0.7
14	Bond market return (1m)	33.3%	70.1%	29.9%	0.4
15	Bond market volatility	32.9%	52.6%	47.4%	0.5
16	Bond market return (12m)	31.9%	12.6%	87.4%	-1.2
17	Bond market order imbalance	31.2%	45.2%	54.8%	-0.3
18	Inflation	28.6%	72.6%	27.4%	0.9
19	Industrial production growth	27.9%	37.0%	63.0%	0.2
20	TED spread (1m)	27.1%	28.6%	71.4%	-0.4
21	Bond order imbalance	25.7%	20.5%	79.5%	-0.2
22	Stock market return (12m)	25.2%	48.8%	51.3%	-0.1
23	TYVIX	24.3%	68.9%	31.1%	0.4
24	VIX	21.9%	69.6%	30.4%	0.0
25	Stock market liquidity	21.2%	67.2%	32.8%	0.2
26	TED spread (6m)	19.8%	63.9%	36.1%	0.4
27	Bond market liquidity	19.5%	69.7%	30.3%	0.4
28	Federal funds rate	18.3%	38.7%	61.3%	0.0

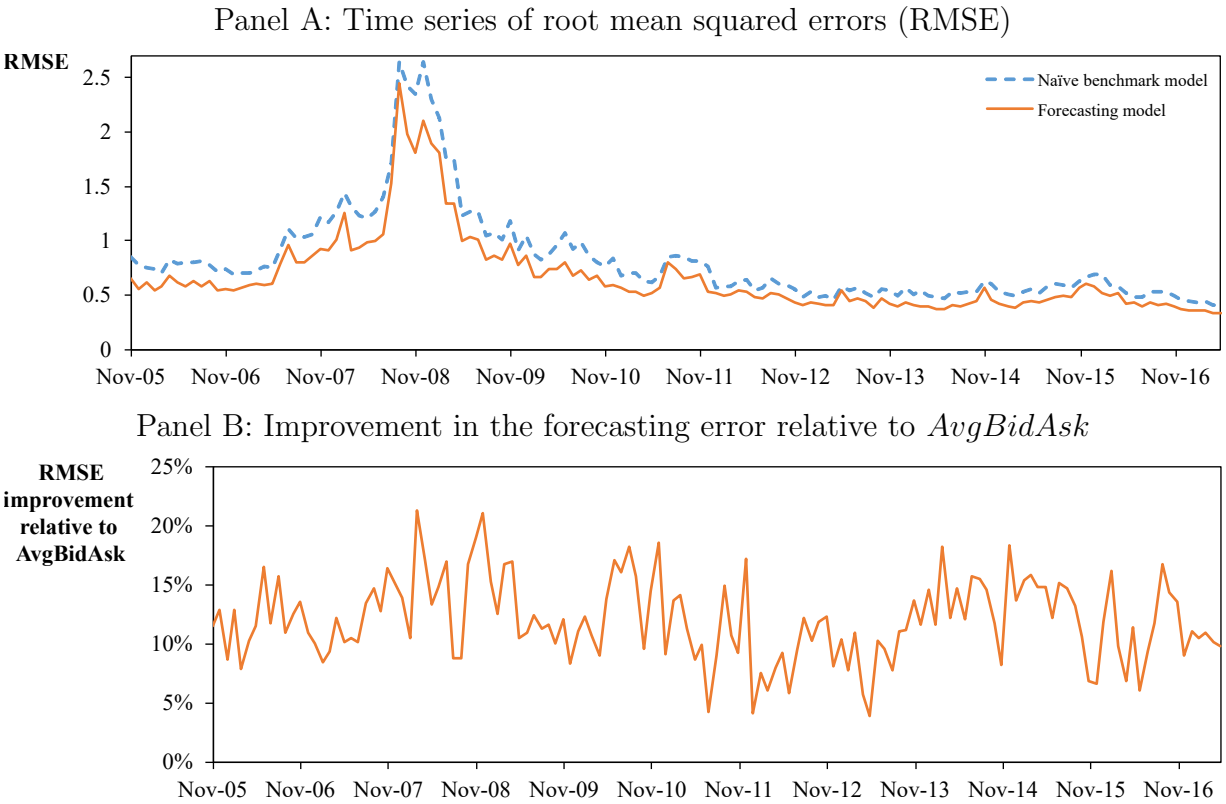


Figure 3.1: **Forecasting performance**
This figure shows the time series performance of the naïve benchmark model and the linear combination model (fc) of Section 3.2.3 for the average bid-ask spread measure *AvgBidAsk*. In Panel A, we report the time series of the monthly root mean squared errors (RMSE). The blue (dashed) line represents errors of the naïve benchmark model and the orange (solid) line the errors of the forecasting model. Panel B shows the improvement in the forecasting error of the forecasting model relative to the average liquidity, i.e., $\frac{RMSE_t^{naïve} - RMSE_t^{fc}}{AvgBidAsk_t}$.

Finally, to investigate the source of our model's forecasting power, we report the predictors that our forecasting algorithm selects and their economic significance in Panel B of Table 3.3. The variables are ranked by their selection frequency, i.e., the percentage of months for which they are included in the model. Because our prediction model is based on three individual selection approaches, we calculate the average frequency. In the same spirit, we report the economic significance as the average across the monthly products of the predictor's average coefficient and its standard deviation in the 12-month rolling window used for calibration. As can be expected, the current liquidity and its 12-month moving average have the strongest impact on next month's forecast. Both variables are almost always included in the predictor set and a liquidity deterioration of one standard deviation is associated with an increase of next month's bid-ask spread of 30.4 and 41.8

basis points, respectively. The variables on rank 3 and 4 are duration and average trade size with an inclusion rate around 90 to 93%. A higher duration is associated with lower liquidity, while larger average trade sizes lead to more narrow spreads. The economic significance of both predictors with 6.7 and -5.3 basis points, is, however, much lower compared to the two autoregressive variables. Consistent with He and Milbradt (2014), we also find credit risk to be a strong driver of future liquidity. A bond's credit spread is in more than 80% of our observation months part of the predictor set and a one standard deviation deterioration in credit quality leads to a 3.9 basis points higher bid-ask spread. On rank 6, the 77% inclusion rate of a bond's segment liquidity emphasizes that a bond's future liquidity is driven by co-movements in the bond market. A one standard deviation decrease in liquidity of a segment is associated with a bid-ask spread increase of 5.1 basis points. Comparing the importance of segment liquidity with that of market-wide liquidity, which is on the second to last rank, indicates that there are systematic differences in the liquidity dynamics between bonds of different characteristics. Finally, total trading volume (rank 7), bid-ask spread volatility (rank 8), amount outstanding (rank 9), and a bond's return (rank 10) complete the set of variables that are selected in more than 50% of the cases.

In the following empirical sections, we evaluate the performance of our liquidity forecast in an asset pricing environment and we examine the effect of expected liquidity on corporate bond fund flows. In such analyses, one faces the challenge that the predictor variables can have an indirect effect on the outcome variable that is not related to liquidity. For example, our candidate predictor set contains information on credit quality, market factors, and macroeconomic variables. To suppress any indirect effects of these variables, we employ a restricted version of our prediction model. To this end, we exclude all variables that are not directly related to liquidity. The restricted set of candidate predictors then only includes the current liquidity, the 12-month moving average of liquidity, volatility of liquidity, average trade size, total trading volume, outstanding amount, segment liquidity, and bond market liquidity.

Since an exclusion of variables with potential predictive power can lead to a loss in forecast accuracy, we check whether the superior performance of the forward-looking approach also holds for the restricted model. Panel A of Table 3.3 shows that the restricted model also outperforms the naïve benchmark model. Because the restricted model requires only a fraction of the input variables, the number of observations is much higher compared to the full model. With 80 basis points, our forecast model generates an error that is 19 basis points lower than the error of the naïve benchmark model. Again, this corresponds

to a relative improvement of about 19%.[43]

3.2.5 Premia for Expected Liquidity in Bond Yields

In addition to the direct evaluation from the previous section, we take an indirect approach to evaluate our forecasting procedure based on an asset pricing analysis. For that, we exploit the prediction of Amihud and Mendelson's (1986) model that illiquid assets command higher expected returns. Because (expected) cash flows to investors depend on future transaction costs, only expected and not current liquidity matters for security prices and expected returns.[44] For that reason, the literature's approach to use today's realized liquidity in asset pricing analyses is identical to employing the naïve proxy for future expected liquidity. In this spirit, Friewald, Jankowitsch, and Subrahmanyam (2012) examine the effect of changes in liquidity on the yield spread of bonds over the Treasury curve. Based on the higher predictive accuracy of our new model compared to the naïve approach, we expect to better capture changes in investors' expectation and, as a result, to better explain the liquidity premium embedded in bond prices.

We test our hypothesis within the setting of Friewald, Jankowitsch, and Subrahmanyam (2012) and perform a monthly panel regression of first differences of yield spreads on changes in expected future liquidity:

$$\begin{aligned}\Delta(\text{Yield spread})_{i,t} &= \alpha + \beta \cdot \Delta(\text{Yield spread})_{i,t-1} + \gamma \cdot \Delta(\widehat{AvgBidAsk})_{i,t+1} \\ &\quad + \delta \cdot \Delta(\text{Controls})_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{3.3}$$

where $\Delta(\widehat{AvgBidAsk})_{i,t+1}$ is the change in predicted liquidity based either on our restricted model of Section 3.2.3 or the naïve approach. Note that $\Delta(\widehat{AvgBidAsk})_{i,t+1}$ contains only information available in month t.[45] Following Friewald, Jankowitsch, and Subrahmanyam (2012), we control for autocorrelation in yield spreads, credit risk, and other liquidity dimensions. The yield spread of a bond is its spread over the Treasury curve calculated

[43] The very similar improvements of the restricted model and the full model over the naïve benchmark are partly due to the additional observations available when using a smaller number of predictors. When we compare the forecasting performance of both models on the same data set, the restricted model is dominated by the full model with a 0.4 basis points lower RMMSE.

[44] Note that in Amihud and Mendelson's (1986) model, transaction costs are constant so that there is no difference between today's realization and future expectations of liquidity. The model can be easily extended to allow for varying transaction costs over time.

[45] For the naïve approach, $\Delta(\widehat{AvgBidAsk})_{i,t+1} = AvgBidAsk_{i,t} - AvgBidAsk_{i,t-1}$ (as $\widehat{AvgBidAsk}^{\text{naïve}}_{i,t+1} = AvgBidAsk_{i,t}$).

via a theoretical bond with the same cash flow structure. We calculate daily yield spreads as volume-weighted average across all trades in TRACE. Bond i's yield spread in month t is then the average across all daily spreads (for more details, see Appendix B.1). We winsorize each month yield spread changes and changes in expected liquidity at the 1% and 99% levels. The credit risk of a bond is represented via changes in 21 rating dummies based on the average rating of the three rating agencies S&P, Moody's, and Fitch. We set the k-th rating dummy to 1 if the average rating is in the interval $[k - 0.5, k + 0.5)$, otherwise we set its value to 0. Finally, to control for other dimensions of liquidity, we employ monthly changes in the logarithm of the outstanding amount, changes in the number of trades, and changes in the logarithm of the average trade size.

The results of the panel regression (3.3) are reported in Table 3.4. Specification (1) only includes the autoregressive term and the control variables and serves as our baseline to evaluate the impact of an inclusion of expected liquidity. We find that yield spreads show a positive autocorrelation that is significant at the 10% level.[46]Consistent with intuition, we find that a higher average trade size is associated with a significantly lower yield spread. For the number of trades, the coefficient is counterintuitively positive. Friewald, Jankowitsch, and Subrahmanyam (2012) attribute this result to investors splitting their trades in an illiquid market. In specification (2), we add a bond's current liquidity as the naïve proxy for expected liquidity to the regression model. Consistent with previous findings, we see a highly significant positive effect. An increase of the bid-ask spread by 1% is associated with a yield spread increase of about 6 basis points. However, the additional explanatory power is rather moderate with an absolute increase in the R^2 of 0.0033 compared to specification (1).[47] If we instead employ our forecasting model in specification (3), we find that the size of the coefficient increases by a factor of about seven. An increase in the expected bid-ask spread of 1% is now associated with an increase in the yield spread of about 45 basis points. The effect on the explanatory power is also more pronounced. The R^2 increases by 0.0184 compared to specification (1), which is more than five times the increase when employing the naïve benchmark.

We test whether this increase in explanatory power can be verified in an out-of-sample setting. To this end, we estimate implied yield spread changes where we employ a backward-looking rolling window of 24 months (with at least 12 months at the start of the

[46]The positive autocorrelation is again consistent with Duffee (1998) and when we switch from a monthly to a weekly frequency, we also get a negative estimate.

[47]The increase in the R^2 is comparable to the results of Friewald, Jankowitsch, and Subrahmanyam (2012). These authors observe an increase of 0.009 when adding four different liquidity measures simultaneously.

Table 3.4: **Yield spread regressions: Expected liquidity proxies**
This table reports results of the panel regression model (3.3) explaining yield spread changes with changes in expected bid-ask spreads. The control variables are the logarithm of the average trade size, the number of trades, and the logarithm of amount outstanding. Further, we employ rating dummies to control for credit risk. We calculate a bond's expected bid-ask spreads for month $t+1$ using either the naïve benchmark or the forecasting model. The naïve forecast for a bond's bid-ask spread in the next month equals the realization of today. The forecasting model is described in Section 3.2.3 and includes only variables that are in the restricted set of candidate predictors, i.e., directly related to liquidity (see Section 3.2.4). We measure bond liquidity using the average bid-ask spread measure. We winsorize changes in yield spreads and changes in expected bid-ask spreads at the 1% and 99% level. Differences in the out-of-sample mean squared errors (MSE) are compared using the test statistic in Harvey, Leybourne, and Newbold (1997). Standard errors are clustered by bond and month and t-statistics are given in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level.

	(1)	(2)	(3)
Intercept	0.0034	0.0044	0.0091
	(0.12)	(0.16)	(0.34)
$\Delta(\text{Yield spread})_{i,t-1}$	0.1760*	0.1749*	0.1667*
	(1.86)	(1.86)	(1.84)
$\Delta(\widehat{AvgBidAsk})^{\text{naïve}}_{i,t+1}$		0.0636***	
		(3.69)	
$\Delta(\widehat{AvgBidAsk})^{\text{fc}}_{i,t+1}$			0.4534***
			(3.49)
$\Delta(\text{Trade size})_{i,t}$	-0.0138***	-0.0061	0.0020
	(-4.08)	(-1.31)	(0.29)
$\Delta(\text{Trades})_{i,t}$	0.0068***	0.0066***	0.0060***
	(3.71)	(3.67)	(3.52)
$\Delta(\text{Amount outstanding})_{i,t}$	0.1347	0.1358	0.1437
	(1.17)	(1.20)	(1.32)
$\Delta(\text{Rating dummies})_{i,t}$	Yes	Yes	Yes
R^2_{adj}	0.0624	0.0657	0.0808
$\Delta(R^2_{adj})$		0.0033	0.0184/0.0151
MSE	0.942	0.939	0.918
Δ(MSE)		-0.003***	-0.024***/-0.021***
		(5.57)	(14.29)/(17.56)
Observations		459,479	

observation period) to calibrate the regression model (3.3). We then calculate the mean squared error between implied and actual changes. To test whether the resulting MSE of specifications (1) to (3) are significantly different, we employ a Diebold and Mariano (1995) test in the spirit of Harvey, Leybourne, and Newbold (1997).[48] Consistent with the in-sample findings, the MSE of 0.942 of the baseline model decreases significantly by 0.003 when adding the current liquidity as the naïve benchmark to the regression. For the specification in which the forecasts are based on our linear combination model, this decrease is much stronger with 0.024. Lastly, we test whether the mean squared errors of the two models are significantly different. And indeed the decrease of 0.021 compared to the benchmark proxy is highly significant. Thus, the new model offers an out-of-sample improvement that is more than eight times the improvement of the naïve model.

Summarizing, our prediction model outperforms the literature's approach to measure expected liquidity using the current liquidity of a bond. We verify the superior performance using two independent analyses. First, we directly compare the predictions with future realizations. Second, our model's forecasts are able to better explain yield spread changes. Notably, the effect of a given bid-ask spread change on bond yield spreads is about seven times larger using the forward-looking approach compared to what is standard in the literature. It is interesting that the 19% lower forecasting error leads to such a strong difference in the relation of expected liquidity with yield spread changes. There are two possible channels that can explain why our forecasts perform so much better compared to realized liquidity. First, much of the variation in liquidity is probably not predictable as it depends on new information becoming available in the next month. Therefore, it is possible that our model captures a much larger part of the variation that is indeed predictable. Second, our bid-ask spread forecast is with a standard deviation of 0.27% for the first differences much less noisy compared to realized liquidity with a standard deviation of 0.84%. Therefore, it is possible that the results using our forecasts are less prone to regression attenuation, which biases coefficients towards zero when regressors are measured with errors.

[48]Note that, as discussed in Section 2.3.1, because of the rolling estimation, the test corresponds to an unconditional Giacomini and White (2006) test, which takes uncertainty in the parameter estimation into account. This test is equally suited to compare the nested and non-nested models of specifications (1) to (3). Moreover, we design in Chapter 4 a test for equal predictive accuracy that controls for the heterogeneity in large panel data. We robustify our results with this newly designed test in Appendix D.

3.3 Fund Flows and Expected Liquidity Deterioration

In this section, we investigate the impact of expected asset liquidity deterioration on corporate bond fund flows. Investors monitor their mutual funds very closely and reallocate money based on previous performance. However, the shape of this flow-performance relation differs between markets. For equity funds, investors reward funds stronger for good performance and are less sensitive to poor performance (see, e.g., Huang, Wei, and Yan (2007)). In contrast, investors in corporate bond funds are rather insensitive to good past performance, but very sensitive to poor past performance. Goldstein, Jiang, and Ng (2017) argue that the concavity of the flow-performance relation for corporate bond funds is related to strategic complementarities among fund investors and the mismatch between the liquidity a fund offers and the illiquidity of its portfolio holdings. If an investor sells her shares, she usually gets the net asset value as of the time of sale. However, portfolio readjustments happen at a later day and liquidation costs then impose negative externalities on the investors who remain in the fund. These effects are stronger when the fund holds more illiquid bonds. Investors take such a first-mover advantage into account and try to preempt other investors when they are considering redeeming the shares of a poorly performing fund.[49] In aggregate, such a behavior can lead to 'runs' on funds similar to bank runs and impair financial stability (see also Chen, Goldstein, and Jian, 2010).

Ultimately, in such a redemption cascade, the first investors get the best outcomes. For that reason, we expect that investors actively try to anticipate liquidity deterioration of poorly performing funds and act on this expectation.

3.3.1 Methodology

We test our hypothesis by examining the flow pattern in corporate bond funds when the liquidity of the fund's assets is expected to decrease. To do so, we select an approach which is inspired by Goldstein, Jiang, and Ng (2017) and perform a monthly panel regression of

[49] In 2016, the SEC adopted a rule that allows funds to adjust their NAVs to reflect liquidation costs (swing pricing). The rule became effective in November 2018 (see also Capponi, Glasserman, and Weber, 2020; Jin, Kacperczyk, Kahraman, and Suntheim, 2019).

flows in corporate bond funds on expected fund asset liquidity changes:

$$\begin{aligned}\text{Flow}_{k,t} =&\beta_0 + \beta_1 \cdot \text{ExpLiqChange}_{k,t+1} + \beta_2 \cdot \text{ExpLiqChange}_{k,t+1} \cdot \mathbb{1}_{\{\text{Alpha}_{k;t-12,t-1}<0\}} \\ &+ \beta_3 \cdot \text{FundLiq}_{k,t} + \beta_4 \cdot \text{FundLiq}_{k,t} \cdot \mathbb{1}_{\{\text{Alpha}_{k;t-12,t-1}<0\}} \\ &+ \beta_5 \cdot \mathbb{1}_{\{\text{Alpha}_{k;t-12,t-1}<0\}} + \gamma \cdot \text{Controls}_{k,t} + \epsilon_{k,t}, \end{aligned} \tag{3.4}$$

where $\text{Flow}_{k,t}$ is the flow of fund k in month t and $\mathbb{1}_{\{\text{Alpha}_{k;t-12,t-1}<0\}}$ is a dummy variable that equals 1 if the past performance of fund k is negative and 0 otherwise. Because there is no reason for investors to anticipate a redemption cascade when past performance is positive, we differentiate between funds with positive and negative performance to analyze the relation of expected liquidity changes and fund flows. Note that the anticipated liquidity change for the month ahead $\text{ExpLiqChange}_{k,t+1}$ can be calculated with information from t. Further, to distinguish between investors who trade on anticipations and those who only use realized liquidity, we include the fund's current asset liquidity $\text{FundLiq}_{k,t}$ as well as an interaction term with past performance. We also include the dummy variable for negative past performance separately and expect that flows are negative when this dummy variable equals 1. Lastly, we follow Goldstein, Jiang, and Ng (2017) and control for a fund's flow in the previous month, total net assets, age, net expense ratio, if redemption fees are charged, and monthly fixed effects.

Monthly flows in individual funds build the basis for our analysis.[50] We calculate the flow for fund k in month t as the relative monthly change in total net assets $\text{TNA}_{k,t}$, adjusted for the fund's return $r_{k,t}$, i.e., $\text{Flow}_{k,t} = \frac{\text{TNA}_{k,t} - \text{TNA}_{k,t-1}(1+r_{k,t})}{\text{TNA}_{k,t-1}}$, using corporate bond fund data from Morningstar (see Appendix B.2 for more details on the data). Following the standard practice in the literature, we winsorize fund flows each month at the 1% and 99% levels. Consistent with Goldstein, Jiang, and Ng (2017), we measure fund liquidity and expected liquidity changes as value-weighted averages across the corporate bonds in the fund's portfolio. Because portfolio holdings are reported at month ends, we use the holdings from the previous month $t-1$ so that the fund flows of the current month cannot influence portfolio compositions, and merge them with asset liquidity in t. For funds that report holdings only quarterly, we use portfolio compositions back to month $t-3$. Liquidity is measured using the average bid-ask spread measure from Section 3.2.1. For expected bond liquidity, we employ the predictions of our restricted forecasting model of

[50] Goldstein, Jiang, and Ng (2017) argue that fund share-level characteristics such as expense ratios, management fees, and redemption fees can have an influence on investor reallocation decisions and thus use individual fund share classes as unit of observation. We follow them and, for ease of readability, use fund and fund share class as synonym for the rest of the chapter.

Section 3.2.3. The expected change in fund k's liquidity for month $t+1$ is then just the relative difference between expected liquidity for $t+1$ and current asset liquidity.

For the separation of funds regarding their past performance, we calculate a fund's average alpha in the preceding twelve months. Again, we follow Goldstein, Jiang, and Ng (2017) and perform time-series regressions of excess fund returns on excess aggregate bond market and stock market returns using a rolling window of the past twelve months. We use the Vanguard Total Bond Market Index Fund return to approximate the aggregate bond market return and the CRSP value-weighted market return for the aggregate stock market return. Fund k's $Alpha_{k;t-12,t-1}$ at month t is then just the estimated intercept of the rolling regression. Finally, for the control variables, we calculate the logarithms of the fund's age and total net assets and create an indicator variable $RearLoad_k$ which equals 1 if the fund charges rear load fees and 0 otherwise.

3.3.2 Results

The results of the panel regression (3.4) are reported in Table 3.5. Before examining the effect of an expected liquidity deterioration, we analyze the relation of flows and the funds' current asset liquidity. To this end, we estimate regression (3.4) without the expected fund liquidity change. This setting is comparable to the original approach of Goldstein, Jiang, and Ng (2017), which targets the amplification of flows out of poorly performing illiquid funds. The main difference here is that we change perspective and set the focus on a fund's asset liquidity rather than on its performance.[51] The results are given in the first specification of Table 3.5. Consistent with Goldstein, Jiang, and Ng (2017), we see a significantly negative coefficient for the interaction of fund asset liquidity and fund performance. An increase of 100 basis points of the average bond bid-ask spread leads to an additional outflow of roughly 0.61% for poorly performing funds. We can interpret this result as a 'run' effect out of struggling funds exacerbated by asset illiquidity. For funds with a positive performance over the last year, we find an insignificant effect on their flows. This finding is consistent with the intuition that positive performance convinces most investors to stay in the fund and thus they are not pressured to sell their shares to avoid the negative externalities of others leaving the fund. On the contrary, it is likely that these funds harvest illiquidity premiums, which might contribute to their good

[51] Goldstein, Jiang, and Ng (2017) also examine the interaction between performance and fund illiquidity for funds with negative past alpha, employing performance as continuous variable and fund liquidity as indicator variable. In the robustness section 3.4.3, we include performance measured via a continuous variable as an additional control.

Table 3.5: **Corporate bond fund flow regression**

This table reports results of the panel regressions of fund flows on expected liquidity changes of Section 3.3.1. We measure a fund's current and expected liquidity as the value-weighted average bid-ask spread across the bonds in the fund's portfolio. Expected liquidity is calculated using the restricted forecasting model of Section 3.2 and expected changes are relative to the current liquidity. Alpha is the intercept of a regression of excess fund returns on excess corporate bond and equity market returns. The dummy variable $\mathbb{1}_{\{\text{Alpha}<\theta\}}$ equals 1 if alpha is below a threshold θ, where we use 0, the upper quartile ($q_{75\%}$), the median ($q_{50\%}$), and the lower quartile ($q_{25\%}$) of all negative alphas. The quartiles correspond to monthly alphas of -5 bps, -12.5 bps, and -30 bps. We include lagged flow, the (natural) logarithm of total net assets, the logarithm of fund age in years, and net expense ratio as controls. We further employ an indicator variable that equals 1 if the fund charges rear load fees and 0 otherwise. The unit of observation is fund share class. We cluster standard errors by fund share class and include month fixed effects. t-statistics and, for cumulated effects, F-statistics are given in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level.

	$\theta = 0$		$\theta = q_{75\%}$		$\theta = q_{50\%}$		$\theta = q_{25\%}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ExpLiqChange		0.0003		0.0004		0.0007		0.0005
		(0.50)		(0.79)		(1.05)		(0.93)
ExpLiqChange × $\mathbb{1}_{\{\text{Alpha}<\theta\}}$		-0.0071***		-0.0083***		-0.0096***		-0.0151***
		(-3.96)		(-4.13)		(-4.28)		(-5.14)
FundLiq	0.0747	0.0289	0.1618	0.1158	0.2278*	0.1890	0.1085	0.0616
	(0.53)	(0.20)	(1.21)	(0.84)	(1.77)	(1.43)	(0.88)	(0.49)
FundLiq × $\mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.6082***	-0.6557***	-0.7053***	-0.7356***	-1.0131***	-1.0848***	-1.2084***	-1.3802***
	(-4.09)	(-4.40)	(-4.63)	(-4.79)	(-6.06)	(-6.41)	(-5.89)	(-6.61)
$\mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.0088***	-0.0080***	-0.0070***	-0.0062***	-0.0035**	-0.0022	0.0001	0.0029
	(-6.41)	(-5.82)	(-4.82)	(-4.24)	(-2.08)	(-1.25)	(0.03)	(1.26)
Lagged flow	0.1514***	0.1513***	0.1517***	0.1516***	0.1519***	0.1518***	0.1525***	0.1523***
	(18.25)	(18.25)	(18.28)	(18.27)	(18.29)	(18.28)	(18.31)	(18.30)
TNA	0.0002	0.0003	0.0003*	0.0003*	0.0004**	0.0004**	0.0004**	0.0004**
	(1.46)	(1.49)	(1.89)	(1.91)	(2.21)	(2.19)	(2.12)	(2.07)
Age	-0.0207***	-0.0207***	-0.0209***	-0.0209***	-0.0211***	-0.0211***	-0.0210***	-0.0210***
	(-30.57)	(-30.60)	(-30.79)	(-30.80)	(-30.96)	(-30.95)	(-30.95)	(-30.94)
Expense	0.0009	0.0010	0.0013	0.0014	0.0009	0.0009	-0.0002	-0.0002
	(0.63)	(0.70)	(0.88)	(0.92)	(0.60)	(0.59)	(-0.12)	(-0.14)
Rear load	-0.0135***	-0.0136***	-0.0137***	-0.0138***	-0.0140***	-0.0140***	-0.0137***	-0.0137***
	(-8.16)	(-8.21)	(-8.30)	(-8.34)	(-8.42)	(-8.43)	(-8.26)	(-8.26)
R^2_{adj}	0.0639	0.0640	0.0637	0.0637	0.0635	0.0635	0.0630	0.0630
Cum. effect ExpLiqChange		-0.0069***		-0.0079***		-0.0090***		-0.0146***
		(15.66)		(16.43)		(17.04)		(25.35)
Cum. effect FundLiq	-0.5335***	-0.6267***	-0.5435***	-0.6198***	-0.7853***	-0.8958***	-1.0999***	-1.3186***
	(16.10)	(21.64)	(14.88)	(18.25)	(24.14)	(29.30)	(30.86)	(41.06)
Econ. sign. ExpLiqChange		-0.16%		-0.19%		-0.26%		-0.45%
Econ. sign. FundLiq	-0.24%	-0.28%	-0.25%	-0.28%	-0.36%	-0.41%	-0.54%	-0.65%
Observations	223,622	223,586	223,622	223,586	223,622	223,586	223,622	223,586

performance. As expected, a negative performance as a standalone dummy variable is associated with a significant outflow. Regarding the control variables, we find that corporate bond fund flows show a significantly positive autocorrelation. Also, older funds experience relatively more outflows. While the fund's total net assets and its net expense ratio do not have a significant impact on its flows, charging rear load fees is, consistent with Goldstein, Jiang, and Ng (2017), significantly associated with flows out of the fund.

Because Goldstein, Jiang, and Ng (2017) show that the run effect is stronger for worse performing funds, we examine three further specifications in which we employ alternative cutoffs for the alpha indicator variable. We use cutoffs that correspond to the upper quartile, the median, and the lower quartile of all negative (monthly) alphas, which are -5 bps, -12.5 bps, and -30 bps, respectively. Across the four specifications (1), (3), (5), and (7) of Table 3.5, we see that the amplification effect is indeed stronger for worse performing funds. While a 100 basis points increase in the average bid-ask spread is associated with a cumulated outflow of 0.53% ($= 0.0747 - 0.6082$, see third-to-last line of Table 3.5) for all negative performing funds in specification (1), this outflow increases to 1.10% when the fund's alpha belongs to the 25% most negative ones (specification (7)). Thus, the cumulated liquidity effect for the lowest alpha quartile is more than doubled compared to the effect for all funds with negative performance.

We now present the results on our main hypothesis that investors anticipate a liquidity deterioration and exit the fund based on their expectation using the full regression model (3.4). In specifications (2), (4), (6), and (8), we always control for the current fund liquidity. This variable captures the costs that investors expect the fund to incur when it has to liquidate assets in response to outflows. Therefore, the effect of an expected liquidity change can be interpreted as investors' attempt to preempt other investors based on expected changes of these costs in the future. While specification (2) of Table 3.5 is based on the indicator variable that simply separates funds according to the sign of their alpha, specifications (4), (6), and (8) use the three different negative performance cutoffs from above. The interaction term between the expected liquidity change and fund performance is negatively significant with t-statistics between 4 and 5 in all specifications. Looking at the cumulated effect, we find that an expected doubling of the bid-ask spread (i.e., an increase of 100%) is associated with significantly stronger outflows of 0.69% for funds with a negative alpha over the last year. As hypothesized, investors seem to leave poorly performing funds in advance if a deterioration in liquidity is expected. For funds with a positive alpha, we see that anticipated changes in liquidity have no effect on flows, which is consistent with the intuition that their investors are not pressured to act strategically. The

cumulated anticipation effect is in all four settings statistically significant. Most importantly, because the incentive for investors to sell their shares as soon as possible increases for worse performing funds, the effect increases monotonously when the performance becomes worse. While an expected 100% increase of the bid-ask spread is associated with an outflow of roughly 0.69% for all funds with a negative alpha, this response increases to 0.90% for the 50% worst performing funds and more than doubles to 1.46% for the 25% worst funds. Consistent with our previous results, we find in each setting a significantly higher flow out of poorly performing funds if their assets are currently more illiquid. Further, we see that the cumulated effect is always statistically significant and again becomes stronger for worse performing funds.

Next, we want to compare the magnitude of the anticipation effect with the effect of realized liquidity. To this end, we calculate the economic significance of both effects for the four different settings at the bottom of Table 3.5. On the one hand, a one standard deviation decrease in (current) fund liquidity is associated with an outflow of 0.28% for funds with negative alpha. This effect increases to 0.65% when the fund's alpha is below -30 bps and thus belongs to the 25% worst funds with negative performance. On the other hand, a one standard deviation of expected liquidity deterioration leads to an outflow of 0.16% for negatively performing funds, which increases to 0.45% for the worst quarter of poor funds. This finding again emphasizes that with decreasing performance, the pressure on investors to redeem their shares immediately is stronger. Comparing both effects, the anticipation effect is roughly 60% to 70% of the size of the realized liquidity effect across the four settings. The comparison with the settings in which we exclude liquidity expectations (specifications (1), (3), (5), and (7)) shows that the economic significance of realized liquidity even slightly increases when including expected liquidity changes. This finding emphasizes that both effects indeed represent separate channels contributing to liquidity-induced 'fund runs'. Thus, only looking at realized liquidity severely underestimates the magnitude of these 'runs' compared to the more complete picture that we provide.

Finally, we want to shed more light on the mechanism behind the anticipation effect. Since strategic complementarities arise for investors only in case of liquidity deterioration, we hypothesize that expected changes in fund liquidity are related asymmetrically to fund flows. To test this hypothesis, we split the anticipated liquidity change $\text{ExpLiqChange}_{k,t+1}$ in equation (3.4) into anticipated liquidity deterioration and improvement. The results are presented in Table 3.6. For poorly performing funds, we observe indeed an asymmetric pattern that only expected liquidity deteriorations have a significant effect on fund flows. Lastly consistent with our previous findings, we see that both expected liquidity

Table 3.6: **Corporate bond fund flow regression - Expected liquidity deterioration vs. improvement**
This table reports results of the panel regressions of fund flows on expected liquidity deterioration and expected liquidity improvement. We measure a fund's current and expected liquidity as the value-weighted average bid-ask spread across the bonds in the fund's portfolio. Expected liquidity is calculated using the restricted forecasting model of Section 3.2 and expected changes are relative to the current liquidity. To distinguish between the effect of expected liquidity deterioration and improvement, we use two dummy variables. $\mathbb{1}_{\{\text{ExpLiqChange}\leq 0\}}$ equals 1 if liquidity is expected to improve and $\mathbb{1}_{\{\text{ExpLiqChange}>0\}}$ is 1 if liquidity is expected to deteriorate. Alpha is the intercept of a regression of excess fund returns on excess corporate bond and equity market returns. The dummy variable $\mathbb{1}_{\{\text{Alpha}<\theta\}}$ equals 1 if alpha is below a threshold θ, where we use 0, the upper quartile ($q_{75\%}$), the median ($q_{50\%}$), and the lower quartile ($q_{25\%}$) of all negative alphas. The quartiles correspond to monthly alphas of -5 bps, -12.5 bps, and -30 bps. We include lagged flow, the (natural) logarithm of total net assets, the logarithm of fund age in years, and net expense ratio as controls. We further employ an indicator variable that equals 1 if the fund charges rear load fees and 0 otherwise. The unit of observation is fund share class. We cluster standard errors by fund share class and include month fixed effects. t-statistics are given in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level.

	$\theta = 0$	$\theta = q_{75\%}$	$\theta = q_{50\%}$	$\theta = q_{25\%}$
ExpLiqChange $\times \mathbb{1}_{\{\text{ExpLiqChange}\leq 0\}}$	-0.0031	-0.0040	-0.0036	-0.0019
	(-0.62)	(-0.80)	(-0.74)	(-0.58)
ExpLiqChange $\times \mathbb{1}_{\{\text{ExpLiqChange}\leq 0\}} \times \mathbb{1}_{\{\text{Alpha}<\theta\}}$	0.0019	0.0056	0.0049	-0.0070
	(0.34)	(1.08)	(0.98)	(-0.48)
ExpLiqChange $\times \mathbb{1}_{\{\text{ExpLiqChange}>0\}}$	0.0004	0.0007	0.0009	0.0007
	(0.80)	(1.07)	(1.25)	(1.16)
ExpLiqChange $\times \mathbb{1}_{\{\text{ExpLiqChange}>0\}} \times \mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.0095***	-0.0127***	-0.0142***	-0.0157***
	(-4.86)	(-5.58)	(-5.65)	(-5.01)
FundLiq	0.0128	0.0923	0.1674	0.0610
	(0.09)	(0.66)	(1.25)	(0.48)
FundLiq $\times \mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.6550***	-0.7336***	-1.0856***	-1.3622***
	(-4.39)	(-4.78)	(-6.43)	(-6.30)
$\mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.0075***	-0.0055***	-0.0013	0.0030
	(-5.44)	(-3.69)	(-0.75)	(1.30)
Controls	Yes	Yes	Yes	Yes
R^2_{adj}	0.0640	0.0637	0.0636	0.0630
Observations	223,586	223,586	223,586	223,586

deterioration and improvement do not have a significant effect on well performing funds.

Summarizing, our results are consistent with investors actively anticipating liquidity deterioration in underperforming funds. Such forecasts will provide them with a first-mover advantage when selling their shares in advance. From the perspective of financial stability, this behavior is dangerous as it could trigger redemption spirals.

3.4 Robustness

In this section, we show that our results are robust against three critical alternative specifications. We first compare our linear forecasting model with a random forest model and show that the empirical results of Sections 3.2.5 and 3.3 are robust also for the random forest. Second, we use a more sophisticated liquidity measure that takes the size dependence of transaction costs into account. Third, we rerun the analyses in Section 3.3 using additional variables and interactions for the funds' previous performance.

3.4.1 Random Forest Prediction Model

In machine learning, the random forest model has become popular as a rather simple non-parametric alternative to classic linear prediction models (see, e.g., Behrens, Pierdzioch, and Risse, 2018; Gu, Kelly, and Xiu, 2020). The basis for this algorithm are regression trees that try to find similar groups among the observations. Simply speaking, the algorithm adds branches to the tree at each step by sorting the observations of the previous node into these new branches based on one of the predictor variables. At the end of this procedure, the terminal nodes (leafs) form the partitions and an observation's predicted value is then the average of the left-hand side variable from all observations in the same partition. Since the predictor variables used for branching as well as their cutoff are chosen to get the minimal forecasting error on the validation set, a single regression tree is prone to overfitting. To overcome this limitation, the random forest model employs a bagging procedure that fits a regression tree for n randomly drawn samples. The final prediction is then simply the average across the predictions of the n regression trees.[52] We employ the random forest procedure with $n = 500$ to our prediction approach of Section 3.2.3 for both the full and the restricted set of candidate predictors. Similar to the out-of-sample

[52] We refer for a more detailed description of the random forest algorithm to Gu, Kelly, and Xiu (2020).

Table 3.7: **Robustness: Forecast performance**
This table reports root mean mean squared errors (RMMSE) for the forecasting model and the naïve benchmark. In Panel A, we employ the random forest model of Section 3.4.1 as forecasting model and measure bond liquidity using the average bid-ask spread measure. The naïve forecast for a bond's liquidity in the next month equals the realization of today. In Panel B, we use the size-adapted average bid-ask spread of Section 3.4.2 to measure a bond's liquidity and calculate expected size-adapted bid-ask spreads employing the forecasting model of Section 3.2.3. The full model includes all candidate predictors from Section 3.2.2, whereas the restricted model only includes variables that are directly related to liquidity (see Section 3.2.4).

Panel A: Random forest model				
	Full model		Restricted model	
	RMMSE	Δ to naïve	RMMSE	Δ to naïve
Forecasting model	0.73	-20.22%	0.80	-19.30%
Naïve benchmark model	0.91		0.99	
Observations	230,790		511,465	
Panel B: Size-adapted liquidity measure				
	Full model		Restricted model	
	RMMSE	Δ to naïve	RMMSE	Δ to naïve
Forecasting model	0.79	-22.21%	0.76	-21.57%
Naïve benchmark model	1.02		0.97	
Observations	230,790		511,465	

cross-validation methods in Section 3.2.3, we use a rolling window to fit the model. In each month t, we split the observations of the previous twelve months into a training set of eleven months and a validation set of the last month.

The predictive accuracy of the random forest model compared to the naïve benchmark is shown in Panel A of Table 3.7. For both the full and the restricted model, we see a decrease in RMMSE of roughly 20%, which is very similar to the improvement of the linear forecasting model in Panel A of Table 3.3. The relation between yield spreads and expected liquidity also remains virtually unchanged if we use the random forest instead of our linear model as a forecasting method (see Panel A of Table 3.8). If anything, the effect of expected liquidity is stronger.

We also analyze the robustness of our results from Section 3.3 that investors consider realized liquidity and anticipated liquidity deteriorations to secure a first-mover advantage when redeeming their corporate bond fund shares. In Table 3.9, we employ the random

Table 3.8: **Robustness: Yield spread regressions**
This table reports results of the panel regression model (3.3) explaining yield spread changes with changes in expected bid-ask spreads for the robustness checks of Sections 3.4.1 and 3.4.2. The control variables are the logarithm of the average trade size, the number of trades, and the logarithm of amount outstanding. Further, we employ rating dummies to control for credit risk. We calculate a bond's expected bid-ask spreads for month $t+1$ using either the naïve benchmark or the forecasting model. The naïve forecast for a bond's liquidity in the next month equals the realization of today. In Panel A, we employ the random forest model of Section 3.4.1 as forecasting model and measure bond liquidity using the average bid-ask spread measure. For convenience, we reprint specifications (1) and (2) from Table 3.4. In Panel B, we use the size-adapted average bid-ask spread of Chapter 2 to measure a bond's liquidity and calculate expected size-adapted bid-ask spreads employing the forecasting model of Section 3.2.3. Again, specification (1) is repeated from Table 3.4. Forecasting models include only variables that are in the restricted set of candidate predictors, i.e., directly related to liquidity (see Section 3.2.4). We winsorize changes in yield spreads and changes in expected (size-adapted) bid-ask spreads at the 1% and 99% level. Differences in the out-of-sample mean squared errors (MSE) are compared using the test statistic in Harvey, Leybourne, and Newbold (1997). Standard errors are clustered by bond and month and t-statistics are given in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level.

Panel A: Random forest model			
	(1)	(2)	(3)
Intercept	0.0034	0.0044	0.0088
	(0.12)	(0.16)	(0.33)
$\Delta(\text{Yield spread})_{i,t-1}$	0.1760*	0.1749*	0.1670*
	(1.86)	(1.86)	(1.85)
$\Delta(\widehat{AvgBidAsk})^{\text{naïve}}_{i,t+1}$		0.0636***	
		(3.69)	
$\Delta(\widehat{AvgBidAsk})^{\text{fc}}_{i,t+1}$			0.4429***
			(3.44)
$\Delta(\text{Trade size})_{i,t}$	-0.0138***	-0.0061	0.0022
	(-4.08)	(-1.31)	(0.32)
$\Delta(\text{Trades})_{i,t}$	0.0068***	0.0066***	0.0059***
	(3.71)	(3.67)	(3.49)
$\Delta(\text{Amount outstanding})_{i,t}$	0.1347	0.1358	0.1737
	(1.17)	(1.20)	(1.56)
$\Delta(\text{Rating dummies})_{i,t}$	Yes	Yes	Yes
R^2_{adj}	0.0624	0.0657	0.0811
$\Delta(R^2_{adj})$		0.0033	0.0187/0.0154
MSE	0.942	0.939	0.915
Δ(MSE)		-0.003***	-0.027***/-0.024***
		(5.57)	(19.62)/(20.04)
Observations		459,479	

Table 3.8 continued

Panel B: Size-adapted liquidity measure			
	(1)	(2)	(3)
Intercept	0.0034	0.0051	0.0155
	(0.12)	(0.19)	(0.62)
$\Delta(\text{Yield spread})_{i,t-1}$	0.1760*	0.1745*	0.1599*
	(1.86)	(1.87)	(1.86)
$\Delta(\widehat{AvgBidAsk})^{\text{naïve}}_{i,t+1}$		0.1112***	
		(3.74)	
$\Delta(\widehat{AvgBidAsk})^{\text{fc}}_{i,t+1}$			0.9300***
			(3.79)
$\Delta(\text{Trade size})_{i,t}$	-0.0138***	-0.0178***	-0.0207***
	(-4.08)	(-5.91)	(-6.84)
$\Delta(\text{Trades})_{i,t}$	0.0068***	0.0066***	0.0060***
	(3.71)	(3.71)	(3.64)
$\Delta(\text{Amount outstanding})_{i,t}$	0.1347	0.1357	0.1458
	(1.17)	(1.21)	(1.40)
$\Delta(\text{Rating dummies})_{i,t}$	Yes	Yes	Yes
R^2_{adj}	0.0624	0.0700	0.1080
$\Delta(R^2_{adj})$		0.0076	0.0456/0.0380
MSE	0.942	0.933	0.882
Δ(MSE)		-0.009***	-0.060***/-0.051***
		(8.23)	(22.80)/(27.40)
Observations		459,479	

forest model to calculate expected liquidity changes. Consistent with our previous findings, we see in specifications (2), (4), (6), and (8) highly significant interaction terms with both realized liquidity and expected liquidity changes. Again, both effects increase in the economic significance from all funds with negative performance (specification (2)) to only funds with the most negative alphas (specification (8)). Quantitatively, the effects are slightly weaker compared to the linear model in Table 3.5.

3.4.2 Size-Adapted Liquidity Measure

Transaction costs in bond markets strongly depend on trade size (see, e.g., Edwards, Harris, and Piwowar, 2007). For that reason, as discussed in Chapter 2, idiosyncratic or systematic variations in the trade size disturb measured transaction costs. Such a bias is

Table 3.9: **Robustness: Corporate bond fund flow regression**

This table reports results of the robustness checks for the impact of expected liquidity deterioration on corporate bond fund flows of Sections 3.4.1, 3.4.2, and 3.4.3. We measure a fund's current and expected liquidity as the value-weighted average across the bonds in the fund's portfolio. In Panel A, we employ the random forest model of Section 3.4.1 as forecasting model and measure bond liquidity using the average bid-ask spread measure. For convenience, specifications (1), (3), (5), and (7) are repeated from Table 3.5. In Panel B, we use the size-adapted average bid-ask spread of Chapter 2 to measure a bond's liquidity and calculate expected size-adapted bid-ask spreads employing the forecasting model of Section 3.2.3. Forecasting models include only variables that are in the restricted set of candidate predictors, i.e., directly related to liquidity (see Section 3.2.4). Expected liquidity changes are relative to the current liquidity. In Panel C, we employ the same specification as in Table 3.5, but include a fund's performance over the last year (alpha) and an interaction term as additional explanatory variables. Alpha is the intercept of a regression of excess fund returns on excess corporate bond and equity market returns. The dummy variable $\mathbb{1}_{\{\text{Alpha}<\theta\}}$ equals 1 if alpha is below a threshold θ, where we use 0, the upper quartile ($q_{75\%}$), the median ($q_{50\%}$), and the lower quartile ($q_{25\%}$) of all negative alphas. The quartiles correspond to monthly alphas of -5 bps, -12.5 bps, and -30 bps. We include lagged flow, the (natural) logarithm of total net assets, the logarithm of fund age in years, and net expense ratio as controls. We further employ an indicator variable that equals 1 if the fund charges rear load fees and 0 otherwise. The unit of observation is fund share class. We cluster standard errors by fund share class and include month fixed effects. t-statistics and, for cumulated effects, F-statistics are given in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level.

Panel A: Random forest model

	$\theta = 0$		$\theta = q_{75\%}$		$\theta = q_{50\%}$		$\theta = q_{25\%}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ExpLiqChange		-0.0001		0.0000		0.0001		0.0001
		(-0.20)		(0.12)		(0.31)		(0.19)
ExpLiqChange $\times \mathbb{1}_{\{\text{Alpha}<\theta\}}$		-0.0043***		-0.0046***		-0.0047***		-0.0066***
		(-3.99)		(-4.13)		(-4.12)		(-5.22)
FundLiq	0.0747	0.0342	0.1618	0.1206	0.2278*	0.1942	0.1085	0.0723
	(0.53)	(0.24)	(1.21)	(0.89)	(1.77)	(1.48)	(0.88)	(0.58)
FundLiq $\times \mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.6082***	-0.6495***	-0.7053***	-0.7223***	-1.0131***	-1.0498***	-1.2084***	-1.2996***
	(-4.09)	(-4.36)	(-4.63)	(-4.71)	(-6.06)	(-6.23)	(-5.89)	(-6.27)
$\mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.0088***	-0.0081***	-0.0070***	-0.0064***	-0.0035**	-0.0027	0.0001	0.0017
	(-6.41)	(-5.88)	(-4.82)	(-4.39)	(-2.08)	(-1.57)	(0.03)	(0.76)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2_{adj}	0.0639	0.0639	0.0637	0.0637	0.0635	0.0635	0.0630	0.0630
Cum. effect ExpLiqChange		-0.0044***		-0.0045***		-0.0046***		-0.0066***
		(18.84)		(18.92)		(18.47)		(28.76)
Cum. effect FundLiq	-0.5335***	-0.6153***	-0.5435***	-0.6017***	-0.7853***	-0.8556***	-1.0999***	-1.2274***
	(16.10)	(21.35)	(14.88)	(17.64)	(24.14)	(27.71)	(30.86)	(37.09)
Econ. sign. ExpLiqChange		-0.15%		-0.17%		-0.20%		-0.35%
Econ. sign. FundLiq	-0.24%	-0.28%	-0.25%	-0.28%	-0.36%	-0.40%	-0.54%	-0.61%
Observations	223,622	223,586	223,622	223,586	223,622	223,586	223,622	223,586

Table 3.9 continued

Panel B: Size-adapted liquidity measure

	$\theta = 0$		$\theta = q_{75\%}$		$\theta = q_{50\%}$		$\theta = q_{25\%}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ExpLiqChange		0.0016		0.0018*		0.0021*		0.0018*
		(1.56)		(1.72)		(1.87)		(1.76)
ExpLiqChange $\times \mathbb{1}_{\{\text{Alpha}<\theta\}}$		-0.0063***		-0.0064***		-0.0066***		-0.0064***
		(-4.58)		(-4.51)		(-4.53)		(-4.43)
FundLiq	0.0002	0.0001	0.0014	0.0013	0.0022*	0.0023*	0.0015	0.0016
	(0.12)	(0.07)	(0.98)	(0.96)	(1.66)	(1.69)	(1.13)	(1.20)
FundLiq $\times \mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.0034**	-0.0039***	-0.0039***	-0.0043***	-0.0055***	-0.0061***	-0.0074***	-0.0083***
	(-2.41)	(-2.73)	(-2.64)	(-2.92)	(-3.19)	(-3.52)	(-3.39)	(-3.75)
$\mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.0110***	-0.0103***	-0.0098***	-0.0090***	-0.0079***	-0.0069***	-0.0047*	-0.0033
	(-8.20)	(-7.67)	(-6.85)	(-6.29)	(-4.46)	(-3.85)	(-1.92)	(-1.36)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2_{adj}	0.0638	0.0639	0.0636	0.0636	0.0634	0.0634	0.0628	0.0629
Cum. effect ExpLiqChange		-0.0048***		-0.0046***		-0.0045***		-0.0047***
		(23.59)		(21.95)		(20.99)		(20.03)
Cum. effect FundLiq	-0.0032**	-0.0038***	-0.0025*	-0.0030**	-0.0033*	-0.0039**	-0.0060***	-0.0067***
	(5.67)	(7.54)	(2.99)	(4.02)	(3.58)	(4.74)	(7.35)	(9.07)
Econ. sign. ExpLiqChange		-0.16%		-0.17%		-0.20%		-0.28%
Econ. sign. FundLiq	-0.16%	-0.18%	-0.12%	-0.15%	-0.16%	-0.19%	-0.31%	-0.34%
Observations	223,622	223,586	223,622	223,586	223,622	223,586	223,622	223,586

Table 3.9 continued

Panel C: Alpha as continuous variable

	$\theta = 0$		$\theta = q_{75\%}$		$\theta = q_{50\%}$		$\theta = q_{25\%}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ExpLiqChange		0.0002		0.0003		0.0005		0.0003
		(0.31)		(0.57)		(0.81)		(0.59)
ExpLiqChange $\times \mathbb{1}_{\{\text{Alpha}<\theta\}}$		-0.0055***		-0.0067***		-0.0083***		-0.0133***
		(-3.15)		(-3.48)		(-3.86)		(-4.65)
FundLiq	-0.0793	-0.1162	-0.0129	-0.0566	0.0422	0.0034	-0.0853	-0.1320
	(-0.55)	(-0.79)	(-0.09)	(-0.40)	(0.32)	(0.03)	(-0.69)	(-1.04)
FundLiq $\times \mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.2270	-0.2773*	-0.3350**	-0.3658**	-0.6854***	-0.7570***	-0.8667***	-1.0383***
	(-1.49)	(-1.81)	(-2.17)	(-2.35)	(-4.10)	(-4.48)	(-4.23)	(-4.98)
$\mathbb{1}_{\{\text{Alpha}<\theta\}}$	-0.0091***	-0.0085***	-0.0071***	-0.0065***	-0.0031*	-0.0020	0.0017	0.0042*
	(-6.67)	(-6.17)	(-4.91)	(-4.45)	(-1.85)	(-1.16)	(0.74)	(1.80)
Alpha	0.5771***	0.5767***	0.7133***	0.7147***	0.8432***	0.8395***	1.1012***	1.0939***
	(3.80)	(3.80)	(4.75)	(4.75)	(5.74)	(5.71)	(7.91)	(7.86)
(Alpha $- \theta$) $\times \mathbb{1}_{\{\text{Alpha}<\theta\}}$	0.4416*	0.3946*	0.1576	0.1125	-0.1883	-0.2331	-0.5482***	-0.6203***
	(1.93)	(1.75)	(0.73)	(0.53)	(-0.96)	(-1.22)	(-2.94)	(-3.44)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2_{adj}	0.0644	0.0644	0.0641	0.0641	0.0639	0.0639	0.0636	0.0637
Cum. effect ExpLiqChange		-0.0053***		-0.0064***		-0.0079***		-0.0130***
		(10.04)		(11.88)		(14.17)		(21.27)
Cum. effect FundLiq	-0.3063**	-0.3935***	-0.3479**	-0.4224***	-0.6432***	-0.7536***	-0.9520***	-1.1703***
	(5.14)	(8.13)	(5.98)	(8.29)	(16.31)	(20.89)	(23.14)	(32.34)
Econ. sign. ExpLiqChange		-0.12%		-0.16%		-0.23%		-0.40%
Econ. sign. FundLiq	-0.14%	-0.18%	-0.16%	-0.19%	-0.30%	-0.35%	-0.47%	-0.58%
Observations	223,622	223,586	223,622	223,586	223,622	223,586	223,622	223,586

particularly strong for those bonds that trade only a few times in a month. In Chapter 2, we develop a procedure to mitigate the related measurement errors. The basic idea is to compare the transaction costs paid for a given volume to the costs usually paid for similar trade sizes. The trade-size adapted liquidity measure is then given as the scaling factor between the observed and the usually paid costs. To rule out that our results are influenced by such measurement errors, we employ daily average bid-ask spreads to estimate monthly scaling factors $sf_{i,t}$ using the model in (2.5). Note that the scaling factor is a relative measure of liquidity. For example, a scaling factor of $sf_{i,t} = 2$ means that bid-ask spreads for bond i in month t are twice as large compared to the average bond in the sample. Therefore, we can use the scaling factors instead of the standard average bid-ask spread measure and apply our forecasting methodology of Section 3.2.3.

The results regarding the predictive accuracy of the forecasting model and the naïve benchmark are presented in Panel B of Table 3.7. Note that the naïve benchmark now assumes that the size-adapted liquidity measure, i.e., the scaling factor is unchanged compared to the previous month. We find a reduction of the RMMSE for this liquidity measure of about 22% for the full and the restricted model compared to the naïve benchmark. The results of our indirect performance evaluation that relies on the relation between changes in yield spreads and changes in expected liquidity are shown in Panel B of Table 3.8. Consistent with the findings in Chapter 2, we find that the modified measure in specification (2) based on the naïve forecast leads to a higher explanatory power and lower out-of-sample error compared to specification (2) in Table 3.4. The explanatory power and out-of-sample error are further improved strongly when moving to differences in expected liquidity calculated with our linear forecasting model in specification (3). The improvements in the R^2_{adj} and the out-of-sample MSE are more than twice as large compared to Table 3.4. One explanation for the stronger improvement is that this measure does not suffer from noise in the estimation due to changing trade-size patterns and as a result can be predicted more precisely. In summary, we also find for this measure that our forward-looking approach directly and indirectly outperforms the naïve benchmark in terms of predictive accuracy.

Next, we employ the size-adapted average bid-ask spread in our fund flow analysis of Section 3.3. The results are provided in Panel B of Table 3.9. For all performance quantiles, we consistently find significantly negative interaction effects of realized liquidity and expected liquidity changes. The impact of both effects on corporate bond fund flows becomes again stronger with decreasing fund performance.

3.4.3 Fund Flows and Alpha

Goldstein, Jiang, and Ng (2017) find that outflows are stronger for more negative alphas. In our main analysis, we capture this dependence with dummy variables for different thresholds of alpha. In Panel C of Table 3.9, we additionally include alpha as a continuous variable and an interaction term with the dummy variable for negative performance. Note that we have to use alpha minus the alpha-threshold in the interaction term so that the functional relation between alpha and flows is continuous. Looking at specification (1), where the dummy variable is one if alpha is negative, we find both a significantly positive coefficient for the continuous alpha and a significantly positive interaction term. This effect is consistent with Goldstein, Jiang, and Ng (2017) and confirms the concavity of fund flows. A response of flows to performance becomes stronger when the alpha is negative. Interestingly, for more negative thresholds, the interaction variable first becomes insignificant and for the most extreme setting, where the dummy variable is only one for the 25% most negative alphas, becomes negative (note that alpha plus the interaction term is still positive). This result indicates that for strongly negative alphas, the relation to flows becomes again flatter.

Most importantly, our results regarding the relation of observed liquidity and expected liquidity changes with the outflows of poorly performing funds remain robust. The cumulated effects and the economic significance again increase with decreasing fund performance.

3.5 Conclusion

In this chapter, we propose a prediction model for individual bond liquidity. Our methodology incorporates a dynamic predictor selection that, first, accounts for the information available to a contemporary forecaster and, second, allows for changing structural relations between the economic variables over time. Our approach is easy to implement for any liquidity measure and we exemplarily apply it to the average bid-ask spread measure.

To assess the performance of the forward-looking approach, we compare it to the literature's current practice to use today's liquidity as the best predictor of future liquidity. We show that our forecasting model outperforms this naïve benchmark in every month of our sample period. The method performs also remarkably well during the financial crisis. Additionally, we perform an indirect performance test exploiting the relation of

changes in market expectations for future liquidity and yield-spread changes. The new prediction model reveals a much stronger dependence of yield spreads on liquidity changes and strongly increases the explainable part of yield spread changes.

Finally, we use our predictions to examine the impact of declining liquidity in poorly performing corporate bond funds on investors' selling decision. Consistent with the implications of strategic complementarities among corporate bond fund investors, we find two types of investor behavior. First, investors actively anticipate liquidity deteriorations in funds with poor past performance and sell their fund shares in advance to secure a first-mover advantage. Second, investors also react stronger for funds with already illiquid portfolio holdings. Both effects become more pronounced for funds with more negative performance during their last twelve months. Our results emphasis the importance of the recent regulatory change that allows to pass on the costs of redemptions to the redeeming shareholders (swing pricing). Because the fire sales from struggling funds might also impact the market as a whole, the regulator should incentivize funds to broadly use this new possibility (for a similar argument, see the theoretical model in Capponi, Glasserman, and Weber, 2020).

Chapter 4

Comparing Forecast Performance of Finance Panel Data Models

4.1 Introduction

With the digitization of financial markets during the last decades, the access for investors and researchers to large data on assets or market participants has tremendously increased. Exploiting this vast font of information to predict financial outcomes for individual assets or market participants becomes more and more popular.[53] In particular, the increasing adoption of machine learning algorithms contributes to this popularity, because their prediction accuracy strongly benefits from the use of extensive panel data due to the algorithms' critical dependence on comprehensive training data. As a result of the large size of the input data, forecasters obtain similarly large panels of predictions. In such large panels of forecasts, however, one usually faces strong heterogeneity in the prediction errors. Thus, given the literature's findings from evaluating panel regressions (see Petersen, 2009), it is mandatory to properly account for the different behavior of individual assets or market participants in the cross-section and across time when comparing the accuracy of different forecast models.

[53]For example, in the stock market, Chinco, Clark-Joseph, and Ye (2019) predict one-minute-ahead returns of individual stocks, while Chen, Hong, and Stein (2001) predict the skewness in their daily returns. Furthermore, Bernoth and Pick (2011) and Liu, Moon, and Schorfheide (2020) employ panel forecast models to predict default probabilities or revenues of individual banks or insurance companies. Note that most of these studies do not test their models for equal predictive accuracy, only Bernoth and Pick (2011) try to draw conclusions from the average of separate Diebold and Mariano (1995) test statistics for each individual.

In this chapter, we introduce new statistical tests to compare the predictive accuracy of competing forecast models for panel data with a vast cross-section and a large time series and respective heterogeneity in the forecast errors. Our main proposed test adaptively controls for the heterogeneity in the data along both dimensions and tests for overall predictive equality in the entire sample. For this, we identify sub-clusters which are homogeneous in their predictive accuracy via unsupervised learning techniques such as hierarchical or k-means clustering in a pre-step. These sub-clusters are then used in a second step to obtain a corrected Diebold and Mariano (1995) type test statistic where the cluster structure crucially determines the estimate of the normalizing standard deviation. We generally show that the clustering matters, i.e., simply ignoring the present heterogeneity can lead to different test results. Moreover, we robustify the pre-step with different clustering techniques and determine how the data-driven shape of the clusters corresponds to economically interpretable subsets. In addition to the overall test, we also suggest a second series of tests for detailed localized insights. For this, we test predictive accuracy in each of the clusters separately and reach an overall decision via multiple testing strategies. Moreover, in this case we exploit pre-knowledge on the clusters from additional observable information in one dimension and the adaptive cluster determination is only required for the second step. With this, we can not only test for overall equal predictive accuracy but identify the driving clusters of the models' forecast performance.

We illustrate the importance of our new tests in the empirical application of Chapter 3 where we predict individual bond liquidity in the U.S. corporate bond market. With a broad spectrum of different bond issuers having various bonds outstanding and the disruptive effects caused by the financial crisis in our sample, we first show that prediction errors of such liquidity forecasts are subject to strong two-dimensional heterogeneity. We then employ our new tests to pairwise compare the predictive accuracy of four different liquidity forecast models. The first model generates simple predictions based on the assumption of a martingale property for liquidity. The three remaining models are based on our dynamic approach of Chapter 3, with each model using a different method to identify liquidity predictors. We find that the dynamic models are significantly more accurate than the simple predictions in various settings. First, all three dynamic models outperform the simple approach regarding the average forecast performance across the entire sample. Furthermore, they also offer a higher precision in each of the time periods before, after, and during the financial crisis as well as when looking separately at investment grade and speculative grade bonds. Among the dynamic forecast models, we mostly cannot reject the null that the different models' forecasts are equally accurate. The only exception are

the two time periods framing the financial crisis where the dynamic model based on an elastic net selection method is relatively more accurate than the other two dynamic models. Interestingly, we find that when we test the nulls of equal predictive accuracy and ignore the panel's heterogeneity or try to compute the loss differential's standard deviation by eliminating the forecasts' panel structure, a large part of the test decisions become either misleading or inconclusive. Without accounting for the forecast errors' heterogeneity, the test statistics are almost always highly inflated and mistakenly lead to a rejection of the null. When exploiting an aggregation of the forecast errors over either time or the cross-section, the decision to reject the null depends in many cases on the choice of the aggregated dimension.

We contribute to the literature on forecast performance tests (Diebold and Mariano, 1995; Giacomini and White, 2006) by studying clustering effects caused by heterogeneity in the standard error of the test statistic. While for standard in-sample significance tests the impact of clustering is well explored under standard assumptions such as a random assignment mechanism (see, e.g., Liang and Zeger, 1986; Petersen, 2009; Abadie, Athey, Imbens, and Wooldridge, 2017), we adaptively determine and exploit the shape of the clusters in an unsupervised pre-step. Moreover, different ways of aggregating forecast deviations along the two panel dimensions lead to different options for test statistics, in this case, with specific advantages depending on the application.

The remainder of this chapter is structured as follows. In Section 4.2, we first introduce our test procedure for the pooled hypothesis test of equal predictive accuracy across the entire sample and second for the multiple hypothesis tests in clusters. We present our empirical setting, the results on the pre-clustering step in this setting as well as the test results for the pooled and joint hypothesis tests in Section 4.3. We robustify our pre-clustering approach in Section 4.4 by employing alternative clustering approaches involving k-means clustering as well as economic reasoning. We conclude in Section 4.5.

4.2 Theory

In this section, we introduce a Diebold and Mariano (1995) type predictive accuracy test in the context of large panel data. Our setting is characterized by large panels, i.e., panels having both an extensive cross-section and a long time series, with respective heterogeneity in each dimension. We assess in a pre-step the two-dimensional heterogeneity in the forecast errors to ensure consistent test results.

4.2.1 Set-up and Test Idea

We work with panel data, thus we observe the realized outcome $y_{i,t+h}$ for each cross-sectional unit $i = 1, \ldots, n$ at time $t + h = 1, \ldots, T$. In our case, T is large, i.e., the forecast period comprises several years of high frequent observations. Also, working with financial data, the cross-section n is vast, for example incorporating several hundred stocks or thousands of bonds. Moreover, we denote by $\hat{y}_{i,t+h|t,m}$ the h-step-ahead forecast of the outcome $y_{i,t+h}$ from model $m = 1, \ldots, M$ conditional on information at time t. For ease of exposition, we formally outline the approach for two competing forecasts $M = 2$ where a large number M can be accommodated by successive pairwise comparisons.

For assessing the predictive performance of different forecasts, we use a loss function $L_{i,t+h|t,m}$ quantifying the accuracy of $\hat{y}_{i,t+h|t,m}$ for $y_{i,t+h}$ (see Diebold and Mariano, 1995). We focus on the empirically most common quadratic loss

$$L_{i,t+h|t,m}(\hat{y}_{i,t+h|t,m}, y_{i,t+h}) = (\hat{y}_{i,t+h|t,m} - y_{i,t+h})^2 = \hat{e}^2_{i,t+h|t,m},$$

where $\hat{e}_{i,t+h|t,m}$ is the squared prediction error SE_m of model m for time $t + h$. The approach can also be easily extended to other types of loss functions (see, e.g., Q-Like loss, Patton, 2011). We abstract from estimation errors in the forecast generation step and take $\hat{y}_{i,t+h|t,m}$ as pseudo observations.

4.2.2 An overall Diebold and Mariano (1995) Test with Pre-clustered Standard Errors

We intend to test for overall equal predictive accuracy of two forecast models m_1 and m_2 across the entire sample. Thus, we work with

$$\bar{L}_m = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} L_{i,t+h|t,m}(\hat{y}_{i,t+h|t,m}, y_{i,t+h})$$

and test for

$$H_0 : E[\bar{L}_{m_1}] = E[\bar{L}_{m_2}] \, . \tag{4.1}$$

For this, we detect systematic differences in the forecast errors

$$\Delta_{i,t+h|t} = \hat{e}^2_{i,t+h|t,m_1} - \hat{e}^2_{i,t+h|t,m_2}$$

with the following pooled Diebold and Mariano (1995) type test statistic

$$\hat{V}_{n,T} = \frac{1}{(nT)} \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\Delta_{i,t+h|t}}{\hat{\sigma}(\Delta_{i,t+h|t})} \tag{4.2}$$

where $\hat{\sigma}(\Delta_{i,t+h|t})$ is an estimate of the standard deviation of $\Delta_{i,t+h|t}$. The difficulty in our setting is to estimate $\sigma(\Delta_{i,t+h|t})$ when the forecast errors are clustered. If this heterogeneity is ignored by using the unadjusted standard deviation and respective estimators, such estimates can be biased, over- or underdetermining the true volatility.[54] Subsequently, the test statistic $V_{n,T}$ can also be biased, resulting in potentially misleading asymptotic and finite sample results of the overall test.

We therefore propose a clustering pre-step for identifying the heterogeneous structure of the standard errors in a data-driven way and suggest a feasible test procedure based on this pre-step.[55] We pre-cluster according to $(\hat{e}_{i,t+h|t,m_1}, \hat{e}_{i,t+h|t,m_2})$ simultaneously in cross-section and time. Thus we work with

$$\hat{\sigma}^2(\Delta_{i,t+h|t}) = \sum_{\kappa=1}^{K} \sum_{\lambda=1}^{L} \frac{|I_{\kappa,\lambda}|}{nT} \hat{\sigma}^2_{\kappa,\lambda}(\Delta_{i,t+h|t}) \tag{4.3}$$

with

$$\hat{\sigma}^2_{\kappa,\lambda}(\Delta_{i,t+h|t}) = \sum_{l=-W}^{W} \sum_{j=-J}^{J} \sum_{i,t \in I_{\kappa,\lambda}} \frac{\left(1 - \frac{l}{M}\right)\left(1 - \frac{j}{J}\right)}{|I_{\kappa,\lambda}|} \left(\Delta_{i-j,t+h-l|t} - \bar{\Delta}\right)\left(\Delta_{i,t+h|t} - \bar{\Delta}\right) \tag{4.4}$$

where (κ, λ) marks each rectangular cluster in cross-section and time with $\kappa = 1, \ldots, K$ and $\lambda = 1, \ldots, L$. With $I_{\kappa,\lambda}$ we count the number of (i, t) for which $(\hat{e}_{i,t+h|t,m_1}, \hat{e}_{i,t+h|t,m_2})$ falls into the cluster (κ, λ) and $|I_{\kappa,\lambda}|$ is the number of observations in $I_{\kappa,\lambda}$. Moreover, $\bar{\Delta}$ is the sample mean of $\Delta_{i,t+h|t}$ in the respective cluster. We think of the clusters as a rectangular partition of the two-dimensional cross-section and time space. Empirically, to partition the two-dimensional space, we employ clustering algorithms with a data-driven choice of the number of clusters. Note that (4.4) is the standard HAC estimator (see Newey and West, 1987) with maximum lag lengths $W > 0$ and $J > 0$ within each cluster and (4.3) pools cluster-specific variances according to the cluster size. Generally, only within clusters, cross-correlations appear while across cluster correlations are negligible.

[54] See for example Petersen (2009) for a detailed overview on the effect of biased standard errors in the context of panel data.

[55] Note that our test procedure is still applicable when the panel data is homogeneous, simply by working with a single rectangular cluster containing all observations.

This is similar to the literature on clustered standard errors for panel data in treatment effects (see, e.g., Liang and Zeger, 1986; Abadie, Athey, Imbens, and Wooldridge, 2017) where, however, cluster assignment generally occurs through some underlying random law and actual realized clusters are not pre-estimated.

For the data-driven detection of clusters, we advocate hierarchical clustering (HC) of the forecast errors $(\hat{e}_{i,t+h|t,m_1}, \hat{e}_{i,t+h|t,m_2})$ separately along the two dimensions.[56] Please see Appendix C.1 for a detailed description of the exact algorithm and the employed cut-off decision for each dimension. In Section 4.4, we robustify these cluster choices by employing alternative clustering approaches involving k-means clustering as well as economic reasoning.

If the detection of clusters is sufficiently precise, then the pooled Diebold and Mariano (1995) test statistic (4.2) asymptotically follows a standard normal distribution, in particular

$$(nT)^{1/2}\hat{V}_{n,T} \xrightarrow{d} N(0,1) \; . \tag{4.5}$$

Note that due to the vast sample size in our empirical setting, the asymptotic distribution very closely approximates the exact distribution of the test statistic in the sample.

4.2.3 Multiple Diebold and Mariano (1995) Tests in Clusters

Instead of the completely data-driven approach in Subsection 4.2.2, we can exploit pre-knowledge about clusters in say, the cross-sectional dimension from additional information regarding e.g. a stock's industry affiliation or different rating classes of bonds (see also Timmermann and Zhu, 2019). This does not only yield efficiency gains but also allows to extract more granular information from separate within-cluster tests. For the joint hypothesis over the entire sample, however, these multiple separate tests then reveal not only if the null is rejected but also where in which parts of the sample it is rejected.

Without loss of generality, we assume that along the cross-sectional dimension the clusters $\tau = 1, \ldots, P$ are set according to additional observable information.[57] We denote the index set of all pairs (i,t) within this pre-set cluster as I_τ. Then only for the remaining time-series dimension the clusters $\lambda = \lambda(\tau) = 1, \ldots, L_\tau$ are determined from data using $(\hat{e}_{i,t+h|t,m_1}, \hat{e}_{i,t+h|t,m_2})$ as a screening device for $i \in I_\tau$. Thus, we work for each cluster τ

[56] We favor hierarchical clustering over top-down clustering algorithms such as k-means because it is less vulnerable to noise in the data and the number of clusters is not pre-determined exogenously (see, e.g., Kaushik and Mathur, 2014).

[57] If time-clusters are pre-set, then results follow immediately by interchanging the role of τ and λ.

with

$$\bar{L}_m^\tau = \frac{1}{|I_\tau|} \sum_{\lambda(\tau)=1}^{L_\tau} \sum_{i,t \in I_{\tau,\lambda(\tau)}} L_{i,t+h|t,m}(\hat{y}_{i,t+h|t,m}, y_{i,t+h}) \tag{4.6}$$

where $(\tau, \lambda(\tau))$ marks each rectangular cluster in cross-section and time with $\tau = 1, \ldots, P$ and the index set $I_{\tau,\lambda(\tau)}$ contains all observation pairs within this cluster. Note, that $\lambda(\tau)$ might vary for different τ and that P does not need to coincide with K in (4.3). Even if κ of (4.3) and τ of (4.6) coincide, the corresponding marginal time-series clusters generally differ.

Thus, within each pre-set cluster τ, we can test for equal predictive accuracy of two forecast models m_1 and m_2:

$$H_0^\tau : E[\bar{L}_{m_1}^\tau] = E[\bar{L}_{m_2}^\tau] \tag{4.7}$$

for all $\tau = 1, \ldots, P$. For this, we detect systematic differences in the loss function $\Delta_{i,t+h|t} = L_{i,t+h|t,m_1}(\hat{y}_{i,t+h|t,m_1}, y_{i,t+h})^2 - L_{i,t+h|t,m_2}(\hat{y}_{i,t+h|t,m_2}, y_{i,t+h})^2 = \hat{e}^2_{i,t+h|t,m_1} - \hat{e}^2_{i,t+h|t,m_2}$ with the following test statistic

$$\hat{V}^\tau = \frac{1}{|I_\tau|} \sum_{i,t \in I_\tau} \frac{\Delta_{i,t+h|t}}{\hat{\sigma}_\tau(\Delta_{i,t+h|t})}. \tag{4.8}$$

and

$$\hat{\sigma}^2_\tau(\Delta_{i,t+h|t}) = \sum_{\lambda(\tau)=1}^{L_\tau} \frac{|I_{\tau,\lambda(\tau)}|}{|I_\tau|} \hat{\sigma}^2_{\tau,\lambda(\tau)}(\Delta_{i,t+h|t})$$

where $\lambda(\tau)$ marks a cluster of the non-clustered dimension in cluster τ with $\lambda(\tau) = 1, \ldots, L_\tau$ and $\hat{\sigma}^2_{\tau,\lambda(\tau)}(\Delta_{i,t+h|t})$ is as in (4.4). If the detection of $\tau = 1, \ldots, P$ is sufficiently precise, we can compute $\hat{V}^\tau$ and obtain the p-value of each τ-specific Diebold and Mariano (1995)-Test from the standard normal distribution of the test statistic analogous to (4.5). Hence we can determine a test decision for each H_0^τ in (4.7).

While these individual cluster-specific results are interesting, our main concern is still an overall test decision on the entire sample. Thus, our focus is in testing the joint hypothesis

$$H_0 = \bigcap_{\tau=1}^{P} H_0^\tau \ . \tag{4.9}$$

For this, we conduct multiple tests with $\hat{V}^\tau$ yielding a separate p-value for each H_0^τ in (4.7) for each cluster $\tau = 1, \ldots, P$. For inferring a decision on the joint hypothesis, we then must adjust the significance level of each individual test according to multiple testing principles. In particular, for controlling the familywise error rate, this requires to set the

test-levels α^τ for the separate clusters in order to achieve an overall test-level α in the joint hypothesis (4.9). We follow the procedure by Holm (1979) which allows for certain dependence in the data.[58] Hence, putting the cluster-specific p-values in ascending order, the νth p-value needs to be assessed against the nominal level

$$\alpha_\nu^\tau = \frac{\alpha}{P - \nu + 1} \, . \tag{4.10}$$

The joint hypothesis (4.9) is then rejected if each individual H_0^τ is rejected for all clusters τ. Thus, for example, to satisfy an overall test-level of $\alpha = 5\%$, $H_0^{\tau_P}$ with the highest p-value must be rejected at the 5%-level, while the νth hypothesis $H_0^{\tau_\nu}$ with a lower p-value must be rejected at the more restrictive $\frac{5}{P-\nu+1}\%$-level. In addition to the overall test result, the joint hypothesis test also locates which clusters and parts of the sample drive the overall result and thus provides additional insights.

As in the previous subsection, we generally propose hierarchical clustering to detect the data-driven marginal partitions $\lambda(\tau)$. Again, we show in Section 4.4 that results are robust with respect to different unsupervised learning approaches.

4.3 Empirical Results

In this section, we compare the forecast performance of several prediction models for individual U.S. corporate bond liquidity between November 2005 and May 2017. The U.S. corporate bond market is an optimal environment to illustrate the importance of our new testing approach. In this market, we observe a large number of issuers with various outstanding bonds and, in addition, disruptive effects in the time series of our sample due to the financial crisis. As a result, individual liquidity forecast errors are likely to exhibit heterogeneity in the cross-section of bonds and across time. We first show that prediction errors are indeed subject to strong two-dimensional heterogeneity. We then employ our new Diebold and Mariano (1995) type tests with pre-clustered standard errors of Sections 4.2.2 and 4.2.3 and show that they are well suited to compare predictive accuracy in large financial forecast panels.

[58] Note that one could also explicitly control for the dependence structure of the test statistics of the joint hypothesis in the spirit of Romano and Wolf (2005) but we found the Holm (1979) scheme sufficient for our empirical applications.

4.3.1 Empirical Setting

For our empirical analyses, we exploit our dynamic forecast approach for individual bond liquidity in the U.S. corporate bond market of Chapter 3. We implement the dynamic panel forecast for the average bid-ask spread measure of Hong and Warga (2000)

$$\text{liquidity}_{i,t} = Avg\left[\frac{\overline{P_{i,d}^{buy}} - \overline{P_{i,d}^{sell}}}{0.5 \cdot \left(\overline{P_{i,d}^{buy}} + \overline{P_{i,d}^{sell}}\right)}\right],$$

where we use bond transaction data from Enhanced TRACE from October 1, 2004 to June 30, 2017 to calculate average daily buy and sell prices $\overline{P_{i,d}^{buy/sell}}$ of bond i for all days d in month t.

The main idea of this approach is that we identify in a first step possible bond liquidity predictors exploiting the large pool of drivers of liquidity such as bond characteristics (e.g., age or credit risk, see Warga, 1992; Mahanti, Nashikkar, Subrahmanyam, Chacko, and Mallik, 2008) or broader influences on bond market liquidity (e.g., spillover effects from the stock market or monetary policy, see Chordia, Sarkar, and Subrahmanyam, 2005; Goyenko and Ukhov, 2009).[59] In the second step, for each month t, we then determine the set of candidate predictors that offer the strongest predictive power during the rolling twelve month calibration period from $t-12$ to t and calibrate the subsequent linear forecast panel model

$$\text{liquidity}_{i,t-k+1} = \alpha_t + \sum_{v \in V_t} \beta_{v,t} \cdot \text{predictor}_{v,i,t-k} + \epsilon_{i,t-k} \tag{4.11}$$

with $k = 1, \ldots, 12$. Consistent with Section 3.2.3, we employ three different methods to determine month t's set of liquidity predictors V_t. The first two methods, a variant of stepwise regression and elastic net, rely on a holdout procedure and select the variable set that leads to the lowest average squared prediction error on the withheld set, the last month of the calibration period. The third method fills the predictor set with the variables that are individually significant at the 5% level during the twelve month calibration period (see also Chudik, Kapetanios, and Pesaran, 2018). Based on each selection method, we then get an estimate for bond i's liquidity $\widehat{\text{liquidity}}_{i,t+1}$ in month $t+1$ using the calibrated values $\widehat{\alpha}_t$ and $\widehat{\beta}_{m,t}$ and the selected predictors' realization in month t.[60] Additionally, we

[59] See Section 3.2.2 and Appendix B.1 for details on the liquidity predictors and on bond transaction data.

[60] In the rare event that $\widehat{\text{liquidity}}_{i,t+1}$ is negative, its value is set to 0. Further, we refer to Section 3.2.3 for the detailed description of the forecast methodology.

Table 4.1: **Descriptive statistics for bond liquidity forecasts**
This table reports descriptive statistics for the panel forecast of individual corporate bond liquidity. Forecasts are generated employing four different models. First, we estimate a bond's future liquidity naïvely assuming a martingale property, i.e., a bond's current liquidity is assumed to be the best estimate for its liquidity in the next month. Second, we use dynamic forecast models based on two holdout procedures or on in-sample significance (see Section 4.3.1 for more details on the forecast models). The forecast period spans from November 2005 to May 2017. Panel A shows general information on the forecast panel data and Panel B the mean squared prediction errors (MSE) for the four competing forecast models.

Panel A: Sample information				
	Observations	Months	Bonds	Avg. bonds per month
	230,790	139	6,462	1,660
Panel B: Forecast performance				
	Naïve model	Holdout - stepwise	Holdout - elastic net	In-sample significance
MSE	0.737	0.495	0.488	0.497

consider the liquidity forecast of a naïve approach that assumes a martingale property for bond liquidity, i.e., $\widehat{\text{liquidity}}_{i,t+1} = \text{liquidity}_{i,t}$.[61] Thus, our final set of competing forecast data consists of four panels based on either the dynamic forecast model (4.11) employing one of the three predictor selection methods or on the naïve approach.

Panel A of Table 4.1 shows general information on the bond liquidity forecasts. Consistent with the setting in Section 4.2, we observe a very large panel of estimates of future bond liquidity. After the initial twelve-month calibration window, our forecast period starts in November 2005 and ends in May 2017 resulting in 139 months in which next month's liquidity is predicted on average for 1,660 bonds. Over the whole observation period, this leads to liquidity predictions for a total of 6,462 bonds and a panel of 230,790 observations as input for the tests for equal predictive accuracy. The mean squared prediction errors of the four competing models are shown in Panel B. We observe the highest MSE of 0.737 for the naïve approach. The three dynamic models have about one third lower MSEs ranging from 0.488 to 0.497 with the elastic net model exhibiting the lowest

[61] In empirical applications, in the absence of an established forecast model, researchers typically have employed a bond's current liquidity when formally its expected value was required (see, e.g., Bao, Pan, and Wang, 2011; Friewald, Jankowitsch, and Subrahmanyam, 2012, among others). As a result, we motivate in Chapter 3 this naïve approach as benchmark for the dynamic forecast models.

average prediction error.

4.3.2 Heterogeneity and Pre-clustering

As discussed in Section 4.2.2, testing for equal precise estimation, i.e., testing the null $H_0 : E[MSE_{m_1}] = E[MSE_{m_2}]$, demands an accurate estimate of the loss differential's standard deviation (4.3). Given the broad variety of corporate bonds and bond issuers in the cross-section as well as the disruptive impact of the financial crisis, we expect however that the models' loss functions exhibit systematic heterogeneity along both dimensions. As supporting evidence, we refer to descriptive stylized facts of the time series and the cross-section of the squared prediction error for the naïve forecast approach.[62] Panel A of Figure 4.1 depicts the distribution of the median squared error across time. We observe a similar pattern for the months before and after the financial crisis with relatively low median squared errors. As expected, the median of the squared error peaks and thus strongly deviates from this pattern during the market turmoil of 2008 and 2009. Additionally, the interquartile range shows an extreme variation in the cross-section during these crisis months and also a considerable cross-sectional variation in the months before and the three years after the financial crisis. Thus, we indeed find consistent with our ex-ante expectation a strong heterogeneity in the forecast errors across time.

Panel B of Figure 4.1 shows the squared error distribution for the cross-section of bonds. For the sake of clarity, we only depict 50 bonds of the five largest issuers with 10 randomly selected bonds for each issuer. Because bonds of different issuers are likely subject to different influences, we find an expected differing behavior for the median squared prediction errors across the selected bonds. While the bonds of AT&T (six-digit CUSIP 00206R) and Walmart (six-digit CUSIP 931142) exhibit rather low median squared errors, the remaining bonds, specifically those of Union Pacific (six-digit CUSIP 907818), are associated with a higher median loss in predictive accuracy as well as with very high interquartile ranges. Again, we find our ex-ante expectation of a strong heterogeneity of the loss function for the cross-section of bonds confirmed. Summarizing, we observe strong heterogeneity for the squared prediction errors along both dimensions, cross-section and time.

Given this strong two-dimensional heterogeneity, we perform the clustering pre-step of

[62] For the sake of brevity, we only report results for the forecast errors based on the naïve approach. Note that, however, the three dynamic models also exhibit similarly strong heterogeneity along the two dimensions.

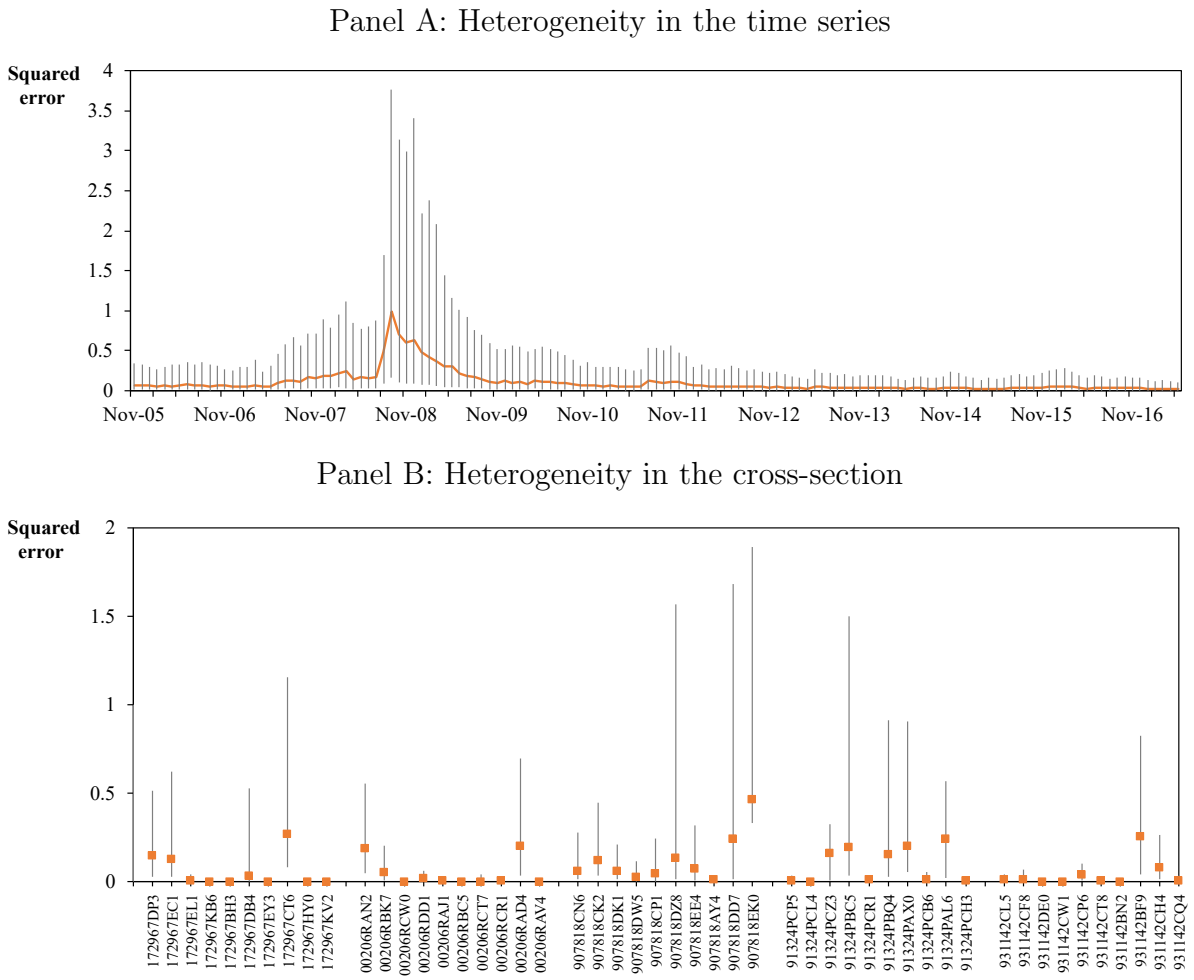


Figure 4.1: **Heterogeneity across time and bonds**
This figure shows the heterogeneity in individual bond liquidity squared prediction errors based on the naïve forecast approach of Section 4.3.1 assuming a bond's current liquidity to be the best estimate for its liquidity in the next month. In Panel A, the orange line depicts the time series of the median squared prediction error. For each month, the vertical gray line indicates the interquartile range. Panel B shows exemplary the heterogeneity of the squared prediction errors across 10 randomly selected bonds from each of the five biggest issuers in the sample, Citigroup (6-digit CUSIPS 172967), AT&T (6-digit CUSIPS 00206R), Union Pacific (6-digit CUSIPS 907818), United Health (6-digit CUSIPS 91324P), and Walmart (6-digit CUSIPS 931142). For each bond, the orange dot depicts the median squared prediction error and the vertical gray line the interquartile range.

Section 4.2.2. As described in Appendix C.1, we first identify for each pairwise comparison of the models of Section 4.3.1 the heterogeneity in the forecast errors in the time dimension of the two competing models by employing hierarchical clustering to their time series of monthly averaged squared prediction errors. We select the optimal cutoff height for the resulting tree as a trade-off between allowing for cluster-within dependencies between a large set of observation months without allowing for spurious correlations. Thus, we obtain for each pairwise comparison the time clusters by cutting the tree at 5% of its height. In the same way, we next identify cross-serial error heterogeneity by employing hierarchical clustering to the cross-section of the average squared prediction errors across time of the two competing models. Again, we select the optimal cutoff height at 1% of the tree's maximum height to satisfy the above trade-off.[63]

Figure 4.2 depicts results of the clustering pre-step. In Panel A, we show the time and cross-sectional clusters for the comparison of the naïve approach and the dynamic model based on stepwise regression. In both models the monthly average squared errors exhibit a similar, almost linear behavior with the crisis months having the highest errors. Accordingly, the hierarchical clustering algorithm groups the neighboring observation months to a set of eight time cluster. For the cross-section of bonds, we also observe similarities between the time series averages of the squared prediction errors of the two models. However, the distribution is much more diverse. Still, our clustering algorithm is able to identify within-cluster similarities and across-cluster heterogeneity with a set of 22 cross-sectional clusters. Panel B, shows the clustering results for the comparison of the two dynamic forecast models based on holdout procedures. Consistent with the strong commonality in the two forecast approaches (only the monthly predictor selection method varies), we observe a strong linear relation between both the time series of monthly average squared prediction errors and the cross-section of the averages of the squared errors across time. As a result, the hierarchical clustering groups subsequent monthly observations or bonds, respectively, into sets of seven time clusters and 20 cross-sectional clusters.[64]

Lastly, we report for all model comparisons descriptive statistics for the rectangular clusters in cross-section and time of the clustering pre-step in Table 4.2. While the time space for the pairwise comparisons is either divided into seven or eight time clusters, the hierarchical clustering identifies 20 to 22 cross-sectional bond clusters. As a result, we

[63]Note that the vast amount of 6,462 bonds results in more diverse average squared errors than for the 139 months. As a result, the Euclidean distance to merge all observations to a single cluster is much higher. Thus, we need to cut at a lower height in order to get a comparable result for the trade-off.

[64]The results of the clustering pre-step for the remaining model comparisons are shown in Appendix C.1.

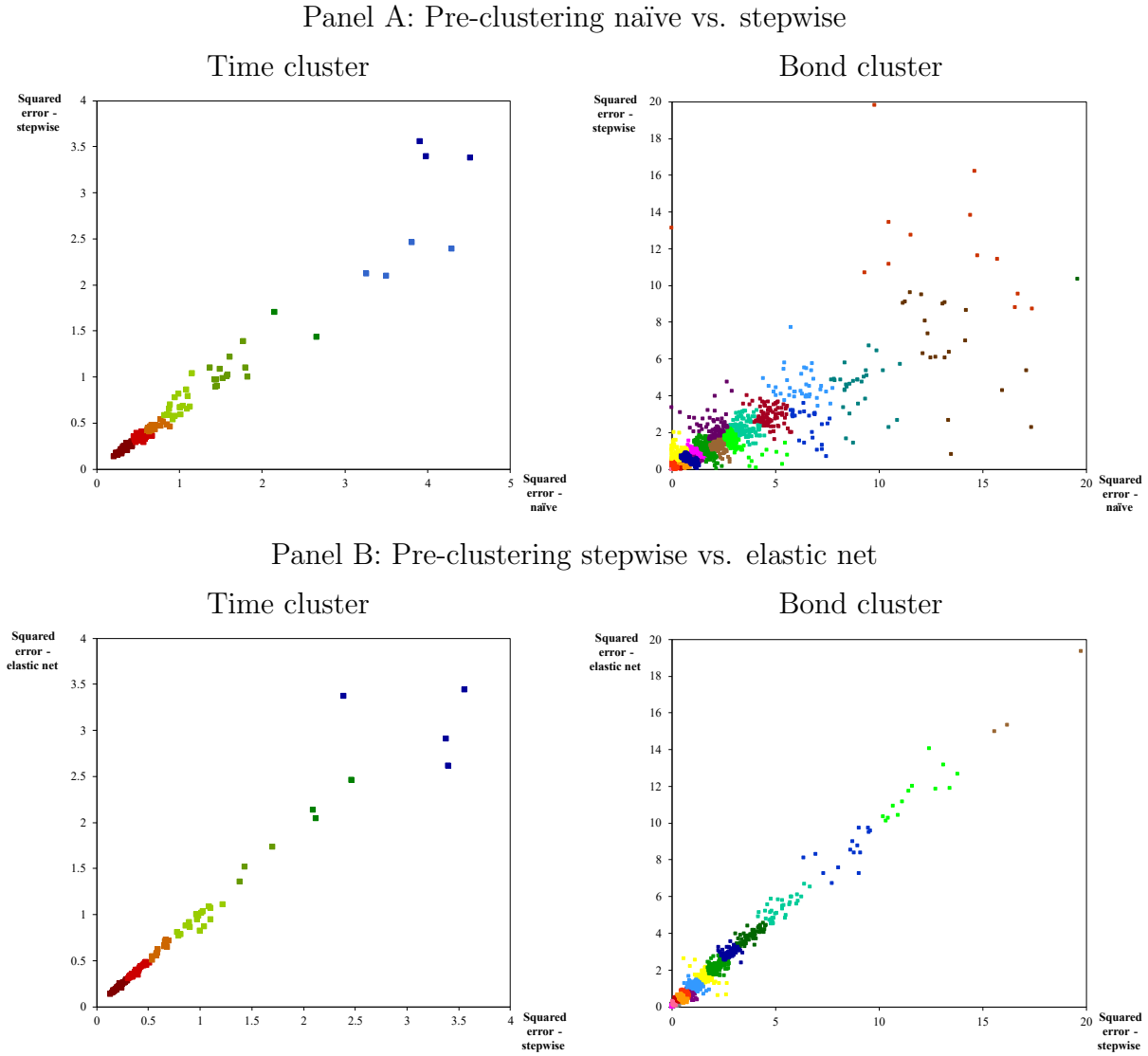


Figure 4.2: **Pre-clustering across time and bonds**
This figure shows exemplary the pre-clustering across time (left image) and bonds (right image) for the comparison of the individual bond liquidity forecast models of Section 4.3.1. Time and cross-sectional clusters are identified by employing hierarchical clustering either on the time series of the average cross-sectional squared prediction errors of the competing models or on the cross-section of the bonds' mean squared prediction error across time. In Panel A, bond liquidity is either predicted by the naïve approach assuming a martingale property for bond liquidity or by the dynamic forecast model based on stepwise regression. In Panel B, both dynamic bond liquidity forecast models exploit a holdout procedure, either stepwise regression or elastic net (for details on the forecast models see Section 4.3.1).

Table 4.2: **Pre-clustering descriptive statistics**
This table reports descriptive statistics of the pre-clustering step for the comparison of the individual bond liquidity forecast models of Section 4.3.1. Bond liquidity is either predicted naïvely assuming a martingale property or employing one of the dynamic forecast models based on either two holdout procedures, stepwise regression and elastic net, or on in-sample significance of Section 4.3.1. Time and cross-sectional clusters are identified by employing hierarchical clustering on either the time series of the average cross-sectional squared prediction errors of the competing models or on the cross-section of the bonds' mean squared prediction error across time. The forecast period spans from November 2005 to May 2017.

	Time	Bond	Observations per cluster pair			
	clusters	clusters	mean	$Q_{25\%}$	$Q_{50\%}$	$Q_{75\%}$
Naïve vs. stepwise	8	22	1334.0	71.0	297.0	1433.0
Naïve vs. elastic net	7	20	1672.4	60.8	367.5	1425.0
Naïve vs. significance	7	21	1602.7	105.0	369.0	1540.8
Stepwise vs. elastic net	7	20	1989.6	80.8	472.5	2216.0
Stepwise vs. significance	8	20	1489.0	92.0	447.0	1794.0
Elastic net vs. significance	8	20	1528.4	106.0	463.0	1825.0

obtain through the intersection of the time and cross-sectional clusters between 116 and 173 two-dimensional rectangular clusters. Each rectangular cluster contains on average a large set of 1,334 to 1,989.6 observations. Even sparsely filled rectangular clusters in the 25%-percentile contain a notable amount of 60.8 to 106 observations. We show in the robustness section 4.4.1 that the size of the rectangular partitions is comparable when employing an alternative unsupervised learning algorithm.

4.3.3 Overall test for equal predictive accuracy

We now want to test the four different forecast models for overall equal predictive accuracy across the entire sample. Thus, we test the pooled null (4.1) with

$$\bar{L}_m = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} L_{i,t+1|t,m} \left(\widehat{\text{liquidity}}_{i,t+1|t,m}, \text{liquidity}_{i,t+1} \right),$$

where we select two of each of the four models m_i and m_j for pairwise comparison.

As discussed, the main difficulty in calculating the pooled Diebold and Mariano (1995) type test statistic (4.2) is to get an accurate estimate of the loss differential's standard

deviation (4.3). Because of the strong heterogeneity in our forecast panel data, this is a particular important task. Thus, we rely on the overall clustered Diebold and Mariano (1995) test of Section 4.2.2 based on the identified two-dimensional clusters of the previous section. In order to examine the impact of the two-dimensional heterogeneity in cross-section and time on the test decisions, we additionally execute the overall Diebold and Mariano (1995) test for four alternative approaches.

In the first test alternative, we do not account for any heterogeneity in the calculation of the loss differential's standard deviation. Consequently, this version serves as a simple baseline from which we can assess the effect of the remaining test alternatives.[65] The next two alternative tests are inspired by Timmermann and Zhu (2019). Based on forecast data of macroeconomic panels, these authors suggest to aggregate the forecast errors in one dimension and compute the standard deviation in the pooled test statistic (4.2) based on the transformed one-dimensional average loss differential. We present the approach when exploiting cross-sectional aggregation. Let $\bar{\Delta}_{t+1|t} = \frac{1}{n}\sum_{i=1}^{n}\Delta_{i,t+1|t}$ and transform the loss differential at $t+1$ to an average loss $R_{t+1|t} = \frac{1}{(n)^{1/2}}\bar{\Delta}_{t+1|t}$. The robust estimate for the standard deviation $\hat{\sigma}(\Delta_{i,t+1|t})$ is then computed by employing a Newey and West (1987) estimator on the time series of $R_{t+1|t}$.[66] Lastly, we compute the standard deviation (4.3) with each bond and each prediction month forming a separate cross-sectional resp. time cluster. As a result, each rectangular cluster in cross-section and time only contains a single observation and we can interpret this approach compared to our data-driven clustering as an upper bound for accounting for the two-dimensional heterogeneity.

The results of the pooled hypothesis test for the five versions are presented in Table 4.3. When comparing the naïve approach with one of the three dynamic forecast models, the difference in the mean squared prediction errors is quite large with 0.240 to 0.249. We see much smaller differences, when we compare the dynamic forecast models among themselves. For the elastic net model that generates on average the lowest forecast errors, we see differences in the MSE of only 0.007 to 0.009 compared to the two remaining forecast models. The difference between the two models based on stepwise regression and on in-sample significance is even smaller with 0.003. As a result, we would expect to reject the null for the pairwise comparisons between the naïve approach and the three dynamic forecast models. Given the small differences among the dynamic models, a rejection of the

[65] Note that while we do not control for the heterogeneity in the two dimensions, we still control for heteroscedasticity by computing White (1984) standard errors.

[66] We calculate the optimal Newey and West (1987) lag length by $l = \left\lfloor \frac{3}{4}T^{1/3} \right\rfloor$. Note that in case of time series aggregation, the approach follows immediately by interchanging the role of t and i and by employing a White (1984) estimator.

Table 4.3: **Overall test for equal predictive accuracy**
This table reports test statistics for the pooled hypothesis test for equal predictive accuracy of Section 4.3.3. We perform pairwise tests for four different forecast models for individual bond liquidity. First, a bond's future liquidity is naïvely predicted assuming a martingale property, i.e., a bond's current liquidity is assumed to be the best estimate for its liquidity in the next month. Second, we employ dynamic forecast models based on two holdout procedures or on in-sample significance (see Section 4.3.1 for more details on the forecast models). Each panel shows differences in the mean squared prediction error and t-statistics in parentheses for each pairwise comparison. In Panel A, we do not make any adjustments for the estimate of the standard deviation in (4.1). In Panel B and C, we employ two procedures inspired by Timmermann and Zhu (2019) where we either exploit an aggregation of the cross-section or alternatively of the time series. In the last two panels, we execute the clustered Diebold and Mariano (1995) test of Section 4.2.2. While each pair of a month and bond marks a separate rectangular cluster in cross-section and time in Panel D, clusters in Panel E are identified data-driven by hierarchical clustering. ** and * indicate statistical significance at the 1% and 5% level.

Panel A: No adjustments

	Naïve model	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	-	0.243**	0.249**	0.240**
		(47.70)	(48.01)	(45.98)
Holdout - stepwise		-	0.007**	-0.003**
			(6.69)	(-2.90)
Holdout - elastic net			-	-0.009**
				(-11.18)

Panel B: Cross-sectional aggregation

	Naïve model	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	-	0.243**	0.249**	0.240**
		(3.23)	(3.87)	(3.97)
Holdout - stepwise		-	0.007	-0.003
			(1.03)	(-0.57)
Holdout - elastic net			-	-0.009
				(-1.33)

Panel C: Time series aggregation

	Naïve model	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	-	0.243**	0.249**	0.240**
		(23.85)	(24.53)	(23.72)
Holdout - stepwise		-	0.007**	-0.003
			(4.56)	(-0.28)
Holdout - elastic net			-	-0.009**
				(-5.45)

Table 4.3 continued

Panel D: Maximum cluster count

	Naïve model	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	-	0.243**	0.249**	0.240**
		(11.86)	(12.23)	(11.60)
Holdout - stepwise		-	0.007	-0.003
			(0.92)	(-0.44)
Holdout - elastic net			-	-0.009
				(-1.29)

Panel E: Hierarchical clustering

	Naïve model	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	-	0.243*	0.249*	0.240*
		(2.43)	(2.36)	(2.38)
Holdout - stepwise		-	0.007	-0.003
			(1.41)	(-0.43)
Holdout - elastic net			-	-0.009
				(-1.38)

null is not obvious.

Looking at the test decisions, we start exemplary with the ex-ante close comparison between the two dynamic forecast models based on stepwise regression and elastic net. In Panel A, our baseline approach in which the covariance matrix is not adjusted for heterogeneity in the two dimensions, we see a very high t-statistic of 6.69 indicating a strong rejection of the null in favor of the elastic net model. In Panels B and C we employ the alternative test versions based on aggregating the loss differential before calculating its standard deviation. These two approaches lead to an inconclusive result. While the null can not be rejected when aggregating prediction errors over the cross-section (t-statistic of 1.03), aggregating the time series is associated with a strong rejection of the null at the 1% level (t-statistic of 4.56). Looking at Table 4.1, we see why these different test outcomes might arise. While we have a large cross-section of on average 1,660 bonds per month, our time series is much smaller with only 139 months. As a result, systematic differences for a part of the bonds may vanish when calculating the average cross-sectional squared prediction error, leading to a trend not to reject the null. On the other hand, when aggregating across time, outliers, e.g., during rare crisis events, can bias the average prediction error

strongly, resulting in a tendency to reject the null. Thus, these aggregation approaches may reduce the computation of the standard deviation to a one-dimensional problem, but at the same time the aggregation of one dimension leads to a loss of potentially important information. Finally, Panel D and E show the results of our approach with pre-clustered standard errors. In both panels, we find low t-statistics of 0.92 and 1.41 indicating an absence of evidence for rejecting the null of equal predictive accuracy. These results are a first indication that without appropriate control for the two-dimensional heterogeneity in the forecast errors we identified in the previous section, test decisions can become misleading or inconclusive. If we ignore the heterogeneity, misleading cross-cluster prediction variance effects lead to a severely underestimated estimate for the standard deviation (4.3) that artificially inflates the t-statistic. Further, due to the information loss when resolving the two-dimensional structure of our large forecast panel, the choice of which dimension is aggregated can have an impact on the test decision.

The remaining comparisons between the dynamic models provide additional support for our preceding findings. Again, unadjusted standard deviations in Panel A mislead to significant rejections at the 1%-level, while the adjusted standard errors in Panel D and E show that the predictor selection method has no significant impact on the overall predictive accuracy. Employing the aggregation approaches, the results of the pre-clustered approach are confirmed for the comparison between the stepwise regression and the in-sample significance models. But for the comparison between the elastic net and the in-sample significance models, the acceptance decision when aggregating the cross-section contradicts the rejection of the null at the 1%-level when aggregating the time series.

Lastly, for the pairwise comparisons of the naïve forecast with the dynamic models, all test approaches reject the hypothesis of equal predictive accuracy in favor of the dynamic models. Interestingly, we still see a similar impact of the test approaches on the test statistic. Unadjusted standard errors lead to highly inflated t-statistics of 45.98 to 48.01 and aggregating the cross-section or the time series leads to either moderate t-statistics of 3.23 to 3.97 or to very high values of 23.72 to 24.53. Finally, comparing the significant t-statistics of the two different pre-clusterings, we observe much higher t-statistics when selecting the smallest possible rectangular clusters across the two dimensions. This indicates that identifying additional cross-serial or time-serial heterogeneity can be important to get reliable test decisions.

Summarizing, our results provide evidence that an adequate consideration of heterogeneity in cross-section and time is imperative when comparing forecast models for overall

equal predictive accuracy and can be achieved with our test procedure of Section 4.2.2. Exploiting this test procedure, we can show that the naïve forecasts are significantly less accurate than those of the three dynamic forecast models, while the choice of the selection method has no significant impact on their overall predictive accuracy.

4.3.4 Multiple tests for equal predictive accuracy

In the preceding section, we have tested the different liquidity forecast models for overall equal predictive accuracy employing a pooled hypothesis test. Alternatively, one could be interested in comparing their forecast performance during specific time periods or for particular classes of bonds. For each time series cluster or cross-sectional cluster, we still face the heterogeneity of the forecast errors in the non-clustered dimension. Thus, we test the joint hypothesis (4.9) of equal predictive accuracy with our test procedure of Section 4.2.3. Again, to examine the test outcomes' dependence on accounting properly for the heterogeneity in the non-clustered dimension, we additionally employ four alternative approaches similar to the previous section.

We exploit pre-knowledge on our forecast panel data and conditional on pre-information from covariates split the panel in either three time clusters or in two cross-sectional bond clusters. Given the disruptive effects of the financial crisis and the following regulatory changes, e.g., by the Volcker rule, we assign observations to either a pre-crisis, crisis, or post-crisis cluster. Based on the NBER recession indicators, observations in the forecast period from November 2005 to December 2007 are assigned to the pre-crisis cluster. The crisis cluster spans from January 2008 to June 2009 and the post-crisis cluster covers the time period from July 2009 to May 2017. In the cross-section, we illustrate our joint hypothesis test with manually selected bond clusters based on a split into investment grade bonds and speculative grade bonds.[67] Each prediction is assigned to the respective bond cluster based on the rating of the underlying bond in the month before the prediction date, i.e., if the bond's average numerical rating across the three rating agencies S&P, Moody's, and Fitch is above 10 (corresponding to BB+), it is assigned to the speculative cluster and otherwise to the investment grade cluster.[68] We test the resulting two joint hypotheses controlling for the familywise error rate to achieve an overall test-level $\alpha = 5\%$

[67] Note that one could also use more sophisticated cross-sectional clusters, for example, based on multilayered portfolio sorts by various bond characteristics. However, for the sake of an easy interpretation of this illustrative example, we use a simple one layer split by rating.

[68] Ratings of S&P and Fitch are transformed to integer value by AAA=1, AA+=2, ... and for Moody's by Aaa=1, Aa1=2, ...

Table 4.4: **Multiple tests in clusters: Descriptive statistics**
This table reports descriptive statistics for the liquidity forecast panel data for manually selected time and cross-sectional bond clusters. We choose three time clusters, prior to the financial crisis, during the financial crisis, and afterwards based on the NBER recession indicators. For cross-sectional clusters, we split bonds into investment grade and speculative grade based on their average rating across the three rating agencies S&P, Moody's, and Fitch. Individual bond liquidity forecasts are generated employing four different models. First, a bond's future liquidity is naïvely predicted assuming a martingale property, i.e., a bond's current liquidity is assumed to be the best estimate for its liquidity in the next month. Second, we use dynamic forecast models based on two holdout procedures or on in-sample significance (see Section 4.3.1 for more details on the forecast models). The forecast period spans from November 2005 to May 2017. Panel A shows general information on the cluster subsamples and Panel B the mean squared prediction errors (MSE) for the four different panel forecast models across the time and cross-sectional cluster subsamples.

Panel A: Sample information

Cluster	Observations	Months	Bonds	Avg. bonds per month
Pre crisis	29,032	26	2,339	1,117
Crisis	22,160	18	2,337	1,231
Post crisis	179,598	95	5,470	1,891
Investment grade	190,370	139	5,657	1,370
Speculative grade	40,420	139	1,416	291

Panel B: Forecast performance

Cluster	Naïve model	Holdout - stepwise	Holdout - elastic net	In-sample significance
Pre crisis	0.717	0.498	0.489	0.530
Crisis	2.559	1.743	1.718	1.735
Post crisis	0.516	0.340	0.336	0.340
Investment grade	0.641	0.430	0.425	0.433
Speculative grade	1.194	0.800	0.785	0.802

and calculate for each cluster the nominal level (4.10) accordingly. Lastly, in case of our test procedure of Section 4.2.3, we identify for each cluster hypothesis the heterogeneity in the non-clustered dimension in a pre-step again via hierarchical clustering.[69]

Sample information on the three time and the two bond cluster subsamples are presented in Panel A of Table 4.4. The post-crisis cluster is the largest among the three time clusters covering roughly 70% of the prediction months and containing about 78% of the observations. Additionally, this cluster has the broadest monthly cross-section with on average 1,891 bonds and 5,470 bonds in total. Nevertheless, both preceding clusters also contain more than 20,000 observations each and exhibit large monthly cross-sections with on average 1,117 to 1,231 bonds. Thus, we expect that the strong heterogeneity in the cross-section is prevalent in all time clusters. In a similar spirit, we see that the investment grade cluster is responsible for the lion's share of the cross-sectional clusters with more than 80% of the observations. But each of the two bond clusters contains observations from the entire forecast period, implying that observations in both clusters are subject to the time series heterogeneity. With respect to the average squared prediction error of the four forecast models of Section 4.3.1 in the time and cross-sectional clusters, we see in Panel B of Table 4.4 a similar pattern compared to the entire sample. The naïve approach always produces by far the highest MSE, while the average errors of the three dynamic models are mostly comparable. Thus, consistent with the pooled hypothesis tests, we expect for both time and cross-sectional clusters to reject the joint null (4.9) when comparing the naïve forecasts with those of one of the three dynamic forecast models. The rejection of the joint null for both cluster variants is analogously not obvious for the pairwise comparisons of the three dynamic forecast models based on different predictor selection methods.

The results of the joint hypothesis tests based on the three manually selected time clusters are presented in Table 4.5. As expected, we find for all approaches that for all three time clusters the p-value is below their critical value of the Holm (1979) scheme when comparing the naïve approach with the dynamic forecast models. Thus, we can reject for these three pairwise comparisons the joint null of equal predictive accuracy at the 5% level in favor of the dynamic models. For the pairwise comparisons of forecast models based on stepwise regression, elastic net, or in-sample significance, we find again differing outcomes among the test alternatives. Without adjusting for the heterogeneity in the cross-section, we find inflated cluster test statistics with p-values virtually being zero when comparing

[69]Similar to Section 4.3.2, we first calculate for the time or cross-sectional cluster's subset of forecast errors the time series averages or the monthly cross-sectional averages, respectively. The dendogram is then again cut at 1% resp. at 5% of the maximum height in case of an underlying time cluster or cross-sectional cluster.

Table 4.5: **Multiple tests for equal predictive accuracy: Time clusters**

This table reports test statistics for the joint hypothesis (4.9) based on three time clusters. Observations are divided into prior to the financial crisis, during the financial crisis, and after the financial crisis based on the NBER recession indicators. We perform pairwise tests for four different forecast models for individual bond liquidity. First, a bond's future liquidity is naïvely predicted assuming a martingale property, i.e., a bond's current liquidity is assumed to be the best estimate for its liquidity in the next month. Second, we employ dynamic forecast models based on two holdout procedures or on in-sample significance (see Section 4.3.1 for more details on the forecast models). Each pairwise comparison entry shows the decision of the joint hypothesis in the first row and afterwards for each time cluster the differences in the mean squared prediction error and the p-value in parentheses. In the first three columns, we do not make any adjustments for the estimate of the standard deviation in the time cluster test statistic (4.8). In the fourth to ninth columns, we test each time cluster employing two procedures inspired by Timmermann and Zhu (2019) where we either exploit an aggregation of the cross-section or alternatively of the time series. In the remaining columns, we execute the joint hypothesis test of Section 4.2.3. While in each time cluster a bond marks a separate cross-sectional cluster in the first three columns, we identify cross-sectional heterogeneity in each time cluster in the last three columns data-driven by hierarchical clustering. Controlling for the familywise error rate to achieve an overall test-level $\alpha = 5\%$, * indicates statistical significance for the respective time cluster with respect to the according critical p-value determined by the Holm (1979) scheme.

		No adjustment			Cross-sectional aggregation			Time series aggregation		
		Holdout - stepwise	Holdout - elastic net	In-sample significance	Holdout - stepwise	Holdout - elastic net	In-sample significance	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	Joint	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0
	Pre crisis	0.219*	0.228*	0.187*	0.219*	0.228*	0.187*	0.219*	0.228*	0.187*
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Crisis	0.817*	0.841*	0.825*	0.817*	0.841*	0.825*	0.817*	0.841*	0.825*
		(0.000)	(0.000)	(0.000)	(0.004)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
	Post crisis	0.176*	0.180*	0.176*	0.176*	0.180*	0.176*	0.176*	0.180*	0.176*
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Holdout - stepwise	Joint		Reject H_0	-		-	-		-	-
	Pre crisis		0.009*	-0.032*		0.009	-0.032		0.009*	-0.032*
			(0.000)	(0.000)		(0.132)	(0.386)		(0.015)	(0.000)
	Crisis		0.024*	0.008		0.024	0.008		0.024	0.008
			(0.012)	(0.341)		(0.673)	(0.743)		(0.894)	(0.364)
	Post crisis		0.004*	0.001		0.004*	0.001		0.004*	0.001*
			(0.000)	(0.129)		(0.017)	(0.668)		(0.000)	(0.019)
Holdout - elastic net	Joint			Reject H_0			-			-
	Pre crisis			-0.041*			-0.041			-0.041*
				(0.000)			(0.273)			(0.000)
	Crisis			-0.016*			-0.016			-0.016
				(0.034)			(0.733)			(0.191)
	Post crisis			-0.003*			-0.003			-0.003
				(0.000)			(0.030)			(0.032)

Table 4.5 continued

		Maximum cluster count			Hierarchical clustering		
		Holdout - stepwise	Holdout - elastic net	In-sample significance	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	Joint	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0
	Pre crisis	0.219*	0.228*	0.187*	0.219*	0.228*	0.187*
		(0.000)	(0.000)	(0.000)	(0.020)	(0.007)	(0.022)
	Crisis	0.817*	0.841*	0.825*	0.817*	0.841*	0.825*
		(0.000)	(0.000)	(0.000)	(0.009)	(0.010)	(0.010)
	Post crisis	0.176*	0.180*	0.176*	0.176*	0.180*	0.176*
		(0.000)	(0.000)	(0.000)	(0.012)	(0.005)	(0.010)
Holdout - stepwise	Joint		Reject H_0	-		-	-
	Pre crisis		0.009*	-0.032*		0.009*	-0.032*
			(0.000)	(0.000)		(0.000)	(0.002)
	Crisis		0.024*	0.008		0.024	0.008
			(0.013)	(0.358)		(0.386)	(0.621)
	Post crisis		0.004*	0.001		0.004*	0.001
			(0.000)	(0.153)		(0.002)	(0.669)
Holdout - elastic net	Joint			Reject H_0			-
	Pre crisis			-0.041*			-0.041*
				(0.000)			(0.000)
	Crisis			-0.016*			-0.016
				(0.031)			(0.289)
	Post crisis			-0.003*			-0.003*
				(0.000)			(0.000)

the elastic net model with the two competing dynamic models. Thus, the joint hypothesis test implies a significant higher predictive accuracy of the elastic net model. In contrast, when exploiting aggregation in the cluster loss function, the two versions inspired by Timmermann and Zhu's (2019) approach suggest an acceptance of the joint null. However, looking at the rejection decisions in the individual clusters, we see that the decision which dimension is aggregated still has a strong impact on the test outcome. While for both comparisons the pre-crisis cluster hypothesis is confirmed when aggregating the cross-section, the same hypothesis is strongly rejected when aggregating across time. Finally, our test procedure of Section 4.2.3 reveals that the predictive accuracy of the three dynamic forecast models is insignificantly different during the financial crisis. As a result, we can not reject the joint hypothesis of equal predictive accuracy across all time clusters. However, outside of the financial crisis, we find a significant higher precision of the elastic net model. Moreover, the stepwise regression model is more accurate compared to the in-sample significance model prior to the financial crisis. Importantly, accounting for the cross-sectional heterogeneity of forecast errors with a maximum number of rectangular clusters leads to a notable underestimation of the loss differential's standard deviation. Thus, our findings confirm the results of Section 4.3.3 and emphasize the importance of a data-driven identification of the heterogeneity in the cross-section.

Lastly, we report the test decisions of the joint hypotheses based on the two manually selected cross-sectional clusters in Table 4.6. We find the same behavior of the alternative test approaches. While inflated cluster test statistics of the baseline approach almost always recommend a rejection of the joint null, the two approaches exploiting an aggregation of the loss differential are inconclusive in their test decisions. Finally, our test procedure controlling for heterogeneity in the non-clustered time dimension by employing hierarchical clustering shows once more that the naïve forecast is dominated by the three dynamic models while no significant different accuracy can be found among them for all cross-sectional clusters. Only for speculative bonds, the dynamic forecast model based on elastic net generates a significant lower prediction error compared to the stepwise regression model.

Table 4.6: **Multiple tests for equal predictive accuracy: Cross-sectional clusters**
This table reports test statistics for the joint hypothesis (4.9) based on two cross-sectional bond clusters. Observations are divided into investment grade bonds and speculative grade bonds based on the underlying bond's average rating across the three rating agencies S&P, Moody's, and Fitch in the month preceding the prediction. We perform pairwise tests for four different forecast models for individual bond liquidity. First, a bond's future liquidity is naïvely predicted assuming a martingale property, i.e., a bond's current liquidity is assumed to be the best estimate for its liquidity in the next month. Second, we employ dynamic forecast models based on two holdout procedures or on in-sample significance (see Section 4.3.1 for more details on the forecast models). Each pairwise comparison entry shows the decision of the joint hypothesis in the first row and afterwards for each cross-sectional cluster the differences in the mean squared prediction error and the p-value in parentheses. In the first three columns, we do not make any adjustments for the estimate of the standard deviation in the cross-sectional cluster test statistic (4.8). In the fourth to ninth columns, we test each cross-sectional cluster employing two procedures inspired by Timmermann and Zhu (2019) where we either exploit an aggregation of the cross-section or alternatively of the time series. In the remaining columns, we execute the joint hypothesis test of Section 4.2.3. While in each cross-sectional cluster a month marks a separate time cluster in the first three columns, we identify time series heterogeneity in each cross-sectional cluster in the last three columns data-driven by hierarchical clustering. Controlling for the familywise error rate to achieve an overall test-level $\alpha = 5\%$, * indicates statistical significance for the respective cross-sectional cluster with respect to the according critical p-value determined by the Holm (1979) scheme.

		No adjustment			Cross-sectional aggregation			Time series aggregation		
		Holdout - stepwise	Holdout - elastic net	In-sample significance	Holdout - stepwise	Holdout - elastic net	In-sample significance	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	Joint	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0
	Investment grade	0.211*	0.216*	0.208*	0.211*	0.216*	0.208*	0.211*	0.216*	0.208*
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Speculative grade	0.394*	0.408*	0.392*	0.394*	0.408*	0.392*	0.394*	0.408*	0.392*
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Holdout - stepwise	Joint		Reject H_0	-		-	-		Reject H_0	-
	Investment grade		0.005*	-0.003*		0.005	-0.003		0.005*	-0.003
			(0.000)	(0.002)		(0.340)	(0.601)		(0.002)	(0.721)
	Speculative grade		0.015*	-0.002		0.015	-0.002		0.015*	-0.002
			(0.000)	(0.517)		(0.229)	(0.694)		(0.001)	(0.436)
Holdout - elastic net	Joint			Reject H_0			-			Reject H_0
	Investment grade			-0.008*			-0.008			-0.008*
				(0.000)			(0.199)			(0.000)
	Speculative grade			-0.017*			-0.017			-0.017*
				(0.000)			(0.267)			(0.001)

Table 4.6 continued

		Maximum cluster count			Hierarchical clustering		
		Holdout - stepwise	Holdout - elastic net	In-sample significance	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve	Joint	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0
model	Investment grade	0.211*	0.216*	0.208*	0.211*	0.216*	0.208*
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Speculative grade	0.394*	0.408*	0.392*	0.394*	0.408*	0.392*
		(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Holdout	Joint		-	-		-	-
- stepwise	Investment grade		0.005	-0.003		0.005	-0.003
			(0.441)	(0.618)		(0.250)	(0.448)
	Speculative grade		0.015	-0.002		0.015*	-0.002
			(0.209)	(0.825)		(0.011)	(0.822)
Holdout	Joint			-			-
- elastic net	Investment grade			-0.008			-0.008
				(0.191)			(0.153)
	Speculative grade			-0.017			-0.017
				(0.259)			(0.173)

4.4 Robustness

In this section, we perform two robustness checks. We first show that results are robust to the choice of the clustering algorithm. To this end, we employ k-means clustering as an alternative to identify cross-sectional and time clusters. Second, we robustify the shape of our data-driven rectangular clusters by showing that our empirical results are comparable when we choose an economically motivated shape for the rectangular clusters.

4.4.1 K-means clustering

In the main analyses, we detect heterogeneity in the two dimensions cross-section and time by employing a bottom-up clustering algorithm. Alternatively, top-down algorithms such as k-means clustering are commonly employed (see, e.g., Leuz, Nanda, and Wysocki, 2003; Fuertes and Kalotychou, 2007; Schmidt, 2015). The main idea of this algorithm is that a user selects a desired number of clusters based on her knowledge on the data, to which all observations closest to the respective cluster center points are assigned. As a first step, the algorithm is initialized by randomly picking for each cluster one observation as center point. Next, the distance to the randomly selected center points is computed for all observations and each observation is assigned to the cluster with the lowest distance to its center point. Based on the classified observations, the center point is then recomputed to represent the center of the underlying cluster. These steps are repeated iteratively until the cluster compositions converge.[70]

We repeat the pooled and joint tests for equal predictive accuracy of Sections 4.3.3 and 4.3.4 employing k-means clustering to the pre-clustering step. For the sake of comparability, we use the same numbers of time and cross-sectional clusters as identified by the hierarchical clustering. Table 4.7 shows descriptive statistics for the two-dimensional rectangular clusters based on k-means clustering. Similar to Table 4.2 in Section 4.3.2, on average, each cluster contains 1,382 to 2,042.4 observations. Regarding the clusters in the 25%-percentile, we see that the k-means algorithm leads in most cases to more sparsely filled rectangular clusters than the hierarchical clustering. Nevertheless, they still comprise a notable amount of at least 39 observations.

The results for the overall clustered Diebold and Mariano (1995) test are presented in

[70] In order to avoid local minima, we execute the k-means clustering for 25 different random starting points and select the cluster run offering the best fit.

Table 4.7: **Robustness: Pre-clustering descriptive statistics**
This table reports descriptive statistics of the pre-clustering step for the comparison of the individual bond liquidity forecast models of Section 4.4.1. Bond liquidity is either predicted naïvely assuming a martingale property or employing one of the dynamic forecast models based on either two holdout procedures, stepwise regression and elastic net, or on in-sample significance of Section 4.3.1. Time and cross-sectional clusters are identified by employing k-means clustering on either the time series of the average cross-sectional squared prediction errors of the competing models or on the cross-section of the bonds' mean squared prediction error across time. The forecast period spans from November 2005 to May 2017.

	Time	Bond	Observations per cluster pair			
	clusters	clusters	mean	$Q_{25\%}$	$Q_{50\%}$	$Q_{75\%}$
Naïve vs. stepwise	8	22	1382.0	43.0	203.0	1458.0
Naïve vs. elastic net	7	20	1735.3	40.0	374.0	1848.0
Naïve vs. significance	7	21	1709.6	39.0	216.0	1138.0
Stepwise vs. elastic net	7	20	2042.4	109.0	451.0	1895.0
Stepwise vs. significance	8	20	1479.4	112.5	429.0	1560.0
Elastic net vs. significance	8	20	1518.4	85.3	430.5	1739.0

Panel A of Table 4.8. For all pairwise comparisons, we find highly comparable t-statistics differing only in the first decimal to the ones based on hierarchical clustering in Panel E of Table 4.3. Moreover, we find in Panel B and C test decisions identical to our previous results in the last three columns of Tables 4.5 and 4.6 for the joint Diebold and Mariano (1995) type tests for the three time clusters and for the two cross-sectional bond clusters. For all pairwise comparisons, the decisions on the joint hypotheses as well as each decision for the individual cluster hypotheses are the same.

4.4.2 Tests with pre-specified clusters

One critique of data-driven clustering is that it can be unclear whether the identified heterogeneity in the cross-section and across time follows any economic reason or is partially based on spurious relations. In general, unsupervised clustering algorithms work fully automatically and are essentially black boxes. As a result, controlling for potentially false heterogeneity could have a severe impact on the pooled and joint test decisions on equal predictive accuracy.

We challenge our detected rectangular clusters identified by hierarchical clustering with pre-specified clusters based on simple economic reasoning. Consistent with our argument

Table 4.8: **Robustness: Test statistics for k-means clustering**
This table reports test statistics for the robustness check of Section 4.4.1 for the pooled and joint hypothesis tests for equal predictive accuracy of Sections 4.3.3 and 4.3.4. Individual bond liquidity is predicted employing four forecast models. First, a bond's future liquidity is naïvely predicted assuming a martingale property, i.e., a bond's current liquidity is assumed to be the best estimate for its liquidity in the next month. Second, we employ dynamic forecast models based on two holdout procedures or on in-sample significance (see Section 4.3.1 for more details on the forecast models). We execute the clustered Diebold and Mariano (1995) tests of Sections 4.2.2 and 4.2.3 with data-driven k-means clusters. In Panel A, we compare the models pairwise for the pooled hypothesis (4.1). In Panel B and C, we test the joint hypothesis (4.9) for either time or cross-sectional clusters (for more details on the selected clusters, see Section 4.3.4). For each pairwise comparison, we show first the difference in the mean squared prediction error. In Panel A, t-statistics are given in parentheses and ** and * indicate significance at the 1% and 5% level. In Panel B and C, p-values are given in parentheses. Controlling for the familywise error rate to achieve an overall test-level $\alpha = 5\%$, * indicates for the two panels statistical significance for the respective cluster with respect to the according critical p-value determined by the Holm (1979) scheme.

Panel A: Pooled hypothesis

	Naïve model	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	-	0.243*	0.249*	0.240*
		(2.32)	(2.24)	(2.19)
Holdout - stepwise		-	0.007	-0.003
			(1.54)	(-0.84)
Holdout - elastic net			-	-0.009
				(-1.41)

Panel B: Joint hypothesis - time clusters

		Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	Joint	Reject H_0	Reject H_0	Reject H_0
	Pre crisis	0.219*	0.228*	0.187*
		(0.022)	(0.008)	(0.022)
	Crisis	0.817*	0.841*	0.825*
		(0.004)	(0.005)	(0.008)
	Post crisis	0.176*	0.180*	0.176*
		(0.016)	(0.014)	(0.023)
Holdout - stepwise	Joint		-	-
	Pre crisis		0.009*	-0.032*
			(0.000)	(0.001)
	Crisis		0.024	0.008
			(0.572)	(0.688)
	Post crisis		0.004*	0.001
			(0.002)	(0.633)
Holdout - elastic net	Joint			-
	Pre crisis			-0.041*
				(0.000)
	Crisis			-0.016
				(0.218)
	Post crisis			-0.003*
				(0.000)

Table 4.8 continued

Panel C: Joint hypothesis - cross-sectional clusters				
		Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	Joint	Reject H_0	Reject H_0	Reject H_0
	Investment grade	0.211*	0.216*	0.208*
		(0.000)	(0.000)	(0.000)
	Speculative grade	0.394*	0.408*	0.392*
		(0.001)	(0.001)	(0.000)
Holdout - stepwise	Joint		-	-
	Investment grade		0.005	-0.003
			(0.556)	(0.448)
	Speculative grade		0.015*	-0.002
			(0.020)	(0.849)
Holdout - elastic net	Joint			-
	Investment grade			-0.008
				(0.123)
	Speculative grade			-0.017
				(0.164)

for the pre-knowledge time clusters in Section 4.3.4, we again cluster observations along the time dimension using the financial crisis as natural cutoff. Thus, we again use the NBER recession indicators and assign observations before 2008, between 2008 and June 2009, and after June 2009 to separate time clusters. For the cross-sectional bond clusters, we make the simple assumption that bonds issued by different companies are subject to a heterogeneous behavior. Thus, we assign bond observations having the same six-digits CUSIP to a cross-sectional cluster. Finally, our rectangular clusters across the two dimensions consist of the observations in the intersection of the time and cross-sectional clusters.

The results of the clustered Diebold and Mariano (1995) tests employing the pre-specified clustering are presented in Table 4.9. For the pooled test statistics in Panel A, we find comparable t-statistics that only differ in one out of the six pairwise comparisons more than by the first decimal with respect to the statistics in Panel E of Table 4.3. For the joint hypothesis based on time clusters selected conditional on pre-information, we find in Panel B identical decisions for 16 of the 18 individual cluster hypotheses of the six pairwise comparisons in the last three columns of Table 4.5. In the same spirit, Panel C shows that 11 of the 12 individual cross-sectional cluster hypotheses share the same test outcome as in the last three columns of Table 4.6. The remaining cluster hypothesis decisions are identical to the outcomes when each month or bond form a separate cluster.[71]

[71]Note that our economically motivated clustering is rather rough and neglects for example possible

Table 4.9: **Robustness: Test statistics for pre-specified clusters**
This table reports test statistics for the robustness check of Section 4.4.2 for the pooled and joint hypothesis tests for equal predictive accuracy of Sections 4.3.3 and 4.3.4. Individual bond liquidity is predicted employing four forecast models. First, a bond's future liquidity is naïvely predicted assuming a martingale property, i.e., a bond's current liquidity is assumed to be the best estimate for its liquidity in the next month. Second, we employ dynamic forecast models based on two holdout procedures or on in-sample significance (see Section 4.3.1 for more details on the forecast models). We execute the clustered Diebold and Mariano (1995) tests of Sections 4.2.2 and 4.2.3 with economically motivated time and bond clusters, i.e., months are clustered prior, during, and after the financial crisis based on the NBER recession indicators while bonds are clustered by issuer. In Panel A, we compare the models pairwise for the pooled hypothesis (4.1). In Panel B and C, we test the joint hypothesis (4.9) for either time or cross-sectional clusters (for more details on the selected clusters, see Section 4.3.4). For each pairwise comparison, we show first the difference in the mean squared prediction error. In Panel A, t-statistics are given in parentheses and ** and * indicate significance at the 1% and 5% level. In Panel B and C, p-values are given in parentheses. Controlling for the familywise error rate to achieve an overall test-level $\alpha = 5\%$, * indicates for the two panels statistical significance for the respective cluster with respect to the according critical p-value determined by the Holm (1979) scheme.

Panel A: Pooled hypothesis

	Naïve model	Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	-	0.243**	0.249**	0.240**
		(2.60)	(2.59)	(2.60)
Holdout - stepwise		-	0.007*	-0.003
			(2.10)	(-0.48)
Holdout - elastic net			-	-0.009
				(-1.24)

Panel B: Joint hypothesis - time clusters

		Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	Joint	Reject H_0	Reject H_0	Reject H_0
	Pre crisis	0.219*	0.228*	0.187*
		(0.000)	(0.000)	(0.000)
	Crisis	0.817*	0.841*	0.825*
		(0.000)	(0.000)	(0.000)
	Post crisis	0.176*	0.180*	0.176*
		(0.000)	(0.000)	(0.000)
Holdout - stepwise	Joint		Reject H_0	-
	Pre crisis		0.009*	-0.032*
			(0.000)	(0.000)
	Crisis		0.024*	0.008
			(0.013)	(0.358)
	Post crisis		0.004*	0.001
			(0.000)	(0.153)
Holdout - elastic net	Joint			Reject H_0
	Pre crisis			-0.041*
				(0.000)
	Crisis			-0.016*
				(0.031)
	Post crisis			-0.003*
				(0.000)

Table 4.9 continued

Panel C: Joint hypothesis - cross-sectional clusters				
		Holdout - stepwise	Holdout - elastic net	In-sample significance
Naïve model	Joint	Reject H_0	Reject H_0	Reject H_0
	Investment grade	0.211*	0.216*	0.208*
		(0.000)	(0.000)	(0.000)
	Speculative grade	0.394*	0.408*	0.392*
		(0.000)	(0.000)	(0.000)
Holdout - stepwise	Joint		-	-
	Investment grade		0.005	-0.003
			(0.441)	(0.618)
	Speculative grade		0.015	-0.002
			(0.209)	(0.825)
Holdout - elastic net	Joint			-
	Investment grade			-0.008
				(0.191)
	Speculative grade			-0.017
				(0.259)

Summarizing, the test results of our pre-clustering based on either economic reasoning or on the hierarchical clustering algorithm are mostly comparable. Thus, our data-driven clustering seems to capture the relevant heterogeneity across time and in the cross-section. Moreover, clustering algorithms do not depend on 'ad-hoc' cluster choices by an econometrician, which may capture all heterogeneity but may also suffer from imprecision.

4.5 Conclusion

In this chapter, we address the problem when panel forecast errors exhibit strong heterogeneity in the presence of a large cross-section and time series. We develop new statistical tests to compare panel forecast models for equal predictive accuracy that specifically account for the heterogeneity in the two dimensions. In particular, our test procedure incorporates a pre-step in which we determine the heterogeneous structure of the forecast errors in a data-driven way by employing hierarchical clustering.

We apply our new testing procedure empirically when comparing the accuracy of individual corporate bond liquidity forecasts of three dynamic prediction models that only

heterogeneity between bonds of different industries or of bonds belonging to different parental companies. Thus, it is likely that the larger data-driven clusters capture some of this additional heterogeneity, leading to the small number of deviating test decisions.

differ in their method of selecting liquidity predictors as well as forecasts of a naïve approach assuming a martingale property for bond liquidity. We show that ignoring or inadequately controlling for the two-dimensional heterogeneity can lead to misleading or inconclusive decisions about rejecting the null of equal predictive accuracy for the pairwise comparisons of the four forecast models. In contrast, our new test procedure shows that while the naïve model is always inferior to the dynamic models, the choice of the predictor selection method has generally no significant impact on the forecast quality of the dynamic approach. Besides the corporate bond market, many financial markets such as the stock market or the option market exhibit a large cross-section and offer data for a long time period. Thus, our results emphasize the importance to control for the two-dimensional heterogeneity whenever comparing financial panel forecast models.

Chapter 5

Summary and Outlook

This dissertation deals with the challenges investors face when measuring and predicting liquidity in bond markets.

Chapter 2 resolves the negative impact of the mechanical link between trading costs and trade volume on standard bond liquidity measures. We introduce a two-stage measurement approach to adapt these measures. This approach eliminates the liquidity estimates' dependence on a bond's average trade size by measuring liquidity relative to a market-wide cost-size function. Consistent with the conceptual superiority, we find ample empirical evidence for the average bid-ask spread measure (see, e.g., Hong and Warga, 2000) and the Schultz (2001) measure that their adapted counterparts are associated with a significant higher liquidity measurement precision. We find that the differences in precision have a strong impact when studying implications of liquidity in empirical applications. While the literature only detects significant premia for the liquidity level and the market liquidity risk in the U.S. corporate bond market when examined separately, we find affirmative results for both sources in a combined Fama-MacBeth regression setup when liquidity is measured accordingly to the unique features of over-the-counter bond trading. As a possibility for future research, we could extent the scope of our study to other over-the-counter markets. For example, as a direct consequence of the financial crisis, the Dodd-Frank Act implemented mandatory post-trade reporting in the U.S. CDS market in December 2012. Moreover, in the quest for transparency on European bond markets, the Markets in Financial Instruments Directive (MiFID II) of the European Securities and Markets Authority entered into force in January 2018 requiring market participants to publish post-trade information for liquid bonds. In both markets, it would be interesting to examine a possible dependence between transaction costs and trade size and test whether

results of the literature (see, e.g., Loon and Zhong, 2016) are robust when this relation is tackled appropriately.

In Chapter 3, we tackle the challenge that investors are currently left alone when forming expectations on future individual bond liquidity. We propose a dynamic forecast approach building on the multifaceted relations of bond liquidity with other bond variables identified by the literature. Exploiting this new approach, we elaborate on the consequences when in empirical applications formally required expected liquidity is proxied by the literature's simple approach to employ current liquidity. We find in several empirical analyses that the impact of investors' expectations is severely underestimated by this approach. Chapter 3 provides potential for at least two additional future research projects. First, from a risk management perspective, it is of high interest for both investors and regulators to be able to prepare a set of countermeasures not only for the average future scenario but also for worst case and best case scenarios. Thus, it might be promising to extend our forecast approach to the possibility of predicting conditional quantiles (see, e.g., Athey, Tibshirani, and Wager, 2019). Second, He and Xiong (2012) and He and Milbradt (2014) argue that a company's probability of default is closely linked to its bonds' secondary market liquidity in a default-liquidity loop. Companies in distress face a deterioration in their bonds' secondary market liquidity. Thus, companies have to rollover their maturing debt by issuing new bonds at a higher face value, subsequently amplifying their financial distress. As a result, it would be interesting to deepen our understanding on the relation between default and liquidity by analyzing the explanatory power of a company's expected bond liquidity at the next rollover date for its current probability of default. Furthermore, it would be interesting to examine whether bond liquidity via the rollover channel or stock liquidity via the information efficiency channel (see Brogaard, Li, and Xia, 2017) provides more information on a company's default.

Chapter 4 provides researchers and practitioners with a tool kit to compare their forecast model in bond markets with a vast cross-section of issuers and their bonds and a long time series. The main challenge for testing for equal predictive accuracy in such a setting is that forecast errors exhibit a heterogeneous behavior across the panel. To solve this problem, we suggest a pre-step that identifies the two-dimensional clusters and a feasible Diebold and Mariano (1995) type test procedure based on this pre-clustering. Our empirical results emphasize the importance to account for the heterogeneity in large panels in order to get reliable and conclusive test decisions. A natural starting point for future research would be to adapt the test procedure to situations when nested forecast models are compared (see, e.g., Clark and McCracken, 2001) or when parameters are subject

to uncertainty in their estimation (see, e.g., Giacomini and White, 2006). Furthermore, it would be interesting to provide a framework in the future that allows forecasters to compare the quality of conditional quantile forecasts of different panel data models (see, e.g., Giacomini and Komunjer, 2005).

Appendix A

Additional Information on *Size-Adapted Bond Liquidity Measures and Their Asset Pricing Implications*

A.1 Data Filters, Bond Yields, and Matching Procedure

We use U.S. corporate bond transaction data from the Enhanced TRACE database from October 1, 2004 to December 31, 2014. Starting with only investment-grade bonds in 2002, TRACE was gradually expanded to finally cover essentially all corporate bonds in October 2004. We apply several data filters to remove bonds with special features and erroneous trade entries. Our filters are similar to the ones used in Schestag, Schuster, and Uhrig-Homburg (2016) and Bongaerts, de Jong, and Driessen (2017). In a first step, we apply the procedures of Dick-Nielsen (2009, 2014) to remove duplicates, withdrawn and corrected entries. Further, we eliminate erroneous entries and extreme outliers by applying the median and reversal filters of Edwards, Harris, and Piwowar (2007). We demand bonds to be actively traded for at least twelve months during our observation period or for at least 50% of the months they are active. Additionally, we exclude observations of defaulted bonds after the default date. We discard perpetuals, convertible and puttable bonds as well as bonds with floating coupon payments. We also demand bonds to be USD denominated,

senior unsecured, and we exclude guaranteed bonds. After applying all filters, our final sample consists of 21,233 bonds and 42,190,265 trades, corresponding to roughly 53% of the about 79 million trades in the U.S. corporate bond market during our observation period.

For the analyses in Sections 2.3.1 and 2.4, we require yields and durations. If the yield-to-maturity is larger than the yield-to-call, the reported yield in TRACE often (but not always) corresponds to the yield-to-call. We calculate both yields and select the one that is closest to the reported yield. We drop observations for which both differences are larger than 1 basis point (about 0.7% of all observations). For the yield spreads in Section 2.3.1, we discount the bond's cash flows with the risk-free Treasury curve to calculate the price of an artificial Treasury bond with the same cash flow structure (using updated data from Gürkaynak, Sack, and Wright (2007) published by the Federal Reserve on http://www.federalreserve.gov). Finally, the yield spread is defined as the continuously compounded yield computed from the reported price minus the corresponding yield calculated from the price of the artificial Treasury bond (see Gehde-Trapp, Schuster, and Uhrig-Homburg, 2018).

For the asset pricing tests in Section 2.4, we calculate expected excess returns using probabilities of default (PD) from the Risk Management Institute (RMI) of the University of Singapore for all publicly traded companies (identified by the stock's ISIN and its Bloomberg Ticker). Since there is no consensus on a procedure to match debt and equity data in the presence of M&A activities,[72] we develop our own approach that makes use of the specific information we get from RMI and Bloomberg. Particularly, RMI provided us with a full list of M&A activities. Moreover, we download from Bloomberg for each bond the 'Issuer Equity' ticker that contains the original issuer of the bond and thus is unchanged after an acquisition. Moreover, we download the 'Bond to Equity' ticker that always references the company currently backing the bond. Using this data, we implement a three-step matching procedure without any manual input. First, we check if the 'Issuer Equity' ticker and the 'Bond to Equity' ticker are identical. If they are, the bond is assigned to this company. Second, if the two tickers differ but we have only PD data either for the original issuer or for the current ticker, we assign the bond to its original or current ticker but only for the time period before the first or back to the last M&A event. In the third step, if the two tickers differ and we have PD data for both underlying companies,

[72]For example, some researchers match bonds and companies via the issuer-specific first 6 digits of the CUSIP and additionally hand-match CUSIPs in case of M&A activities (see, e.g., Feldhütter and Schaefer, 2018). Others use 6-digit CUSIPS and historical CUSIPS in CRSP (see, e.g., Chordia, Goyal, Nozawa, Subrahmanyam, and Tong, 2017 or Ederington, Guan, and Yang, 2015).

we check if we can track the acquisition path in RMI's M&A list and assign this path if possible. If we find only an incomplete path, we try to complete it with corporate action data from Bloomberg by checking if the last available company on the unfinished path is a target/parent of the 'Bond to Equity' ticker.[73] Finally, if the 'Issuer Equity' ticker is not listed in RMI's M&A list, we use Bloomberg's corporate action functionality and check if the 'Bond to Equity' ticker is the acquirer/spin-off of the 'Issuer Equity' ticker and assign this direct path if possible. In total, our matching procedure is able to assign company-specific PDs to 16,742 bonds in the sample.

A.2 Information Content of Small and Large Trades

An econometrician who wants to measure the liquidity of a bond using past transactions has to decide on the data that provide the most accurate and comprehensive information. A large strand of the literature (see, e.g., Dick-Nielsen, Feldhütter, and Lando, 2012 or Feldhütter, Hotchkiss, and Oğuzhan, 2016) exclude retail trades below \$100,000 for the calculation of liquidity measures, essentially assuming that small trades provide no meaningful information. Because about two thirds of all transactions in the corporate bond market are below \$100,000, it is often not possible to calculate a daily or even monthly liquidity measure just from large trades. Panel A of Table A.1 shows that for the common Schultz (2001) measure, the number of months for which a bond's liquidity can be assessed drops by roughly 22% when ignoring small trades below \$100,000. For the average bid-ask spread measure, this drop is with 42% even more pronounced as this measure has more stringent data requirements (both standard measures are explained in Sections 2.2.2 and 2.2.3). On a daily frequency, the problem is exacerbated and the observations for which a liquidity measure can be calculated are reduced by more than 50% for both measures when using only large trades.

We next challenge the literature's implicit assumption that small trades are uninformative for institutional trading costs when large trades are available. To this end, we run a panel regression explaining a bond's daily large-trade transaction costs with monthly bid-ask spread estimates based on large or small trades:

$$tc_{i,d,t}^{\text{large}} = \alpha + \beta^{\text{large}} \cdot tc_{i,t\setminus\{d\}}^{\text{large}} + \beta^{\text{small}} \cdot tc_{i,t\setminus\{d\}}^{\text{small}} + \epsilon_{i,d,t}, \tag{A.1}$$

[73] If we are not able to complete the acquisition path, we partly assign the bond to the tickers for which we know the mapping and drop it for the remaining period.

Appendix A. Additional Information on Size-Adapted Bond Liquidity Measures and Their Asset Pricing Implications

Table A.1: **Information content of small and large trades**

Trades with a notional < \$100,000 are classified as small trades, volumes ≥ \$100,000 are large trades. We measure transaction costs with either the Schultz (2001) measure or the average bid-ask spread measure. In Panel A, we report for a monthly and daily frequency the number of observations for which a transaction cost measure can be calculated using either all trades or only large trades. Panels B and C report results for a panel regression explaining daily transaction costs of large or small trades with monthly transaction costs based on either large or small trades:

$$tc_{i,d,t}^{\text{large/small}} = \alpha + \beta^{\text{large}} \cdot tc_{i,t\setminus\{d\}}^{\text{large}} + \beta^{\text{small}} \cdot tc_{i,t\setminus\{d\}}^{\text{small}} + \epsilon_{i,d,t}.$$

For the monthly measures, we use information from all days in the month excluding the day under consideration. In Panel B (C), we estimate the regression models with daily transaction costs of large (small) trades on the left-hand side. In specifications (1) and (4) ((2) and (5)), we include only monthly transaction costs of large (small) trades on the right-hand side. Specifications (3) and (6) combine both categories. Standard errors are clustered by bond. The t-statistics are given in parentheses. ** and * represent statistical significance at the 1% and 5% level.

Panel A: Observations with available liquidity measure

	Daily		Monthly	
	Schultz (2001)	Avg. bid-ask spread	Schultz (2001)	Avg. bid-ask spread
All trades	2,967,320	2,726,414	388,029	523,096
Large trades only	1,435,389	1,210,404	302,904	303,319

Panel B: Explaining transaction costs of large trades

	Schultz (2001)			Average bid-ask spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0025**	0.0012**	0.0008**	0.0012**	0.0005**	0.0002**
	(43.59)	(15.79)	(18.05)	(33.75)	(7.24)	(6.40)
$tc_{i,t\setminus\{d\}}^{\text{large}}$	0.6257**		0.4548**	0.7895**		0.6062**
	(70.57)		(49.86)	(123.15)		(65.07)
$tc_{i,t\setminus\{d\}}^{\text{small}}$		0.3434**	0.1792**		0.3607**	0.1452**
		(47.55)	(42.13)		(45.60)	(30.81)
R_{adj}^2	0.2300	0.1829	0.2626	0.3391	0.2514	0.3616
Observations		742,159			358,740	

Panel C: Explaining transaction costs of small trades

	Schultz (2001)			Average bid-ask spread		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0092**	0.0026**	0.0025**	0.0070**	0.0013**	0.0012**
	(60.24)	(27.58)	(26.72)	(46.04)	(28.49)	(27.45)
$tc_{i,t\setminus\{d\}}^{\text{large}}$	0.8623**		0.1763**	1.2458**		0.1881**
	(57.27)		(20.76)	(68.22)		(25.27)
$tc_{i,t\setminus\{d\}}^{\text{small}}$		0.7832**	0.7196**		0.9047**	0.8379**
		(119.35)	(89.08)		(261.07)	(157.74)
R_{adj}^2	0.1987	0.4329	0.4383	0.3185	0.5965	0.6005
Observations		742,159			358,740	

where we calculate the monthly measures $tc_{i,t\setminus\{d\}}$ for bond i in month t excluding the day under consideration d. A significant estimate for β^{small} would show that small trades help to explain the transaction costs of large trades. The results of the panel regressions are presented in Panel B of Table A.1. In specifications (1) and (4), we explain transaction costs of large trades using only the monthly estimates from large transactions, whereas in specifications (2) and (5), we only employ small trades. Finally, specifications (3) and (6) use both size categories. As expected, large trades from other days of the same month carry information for a bond's institutional transaction costs with an R^2_{adj} of 23% for the Schultz (2001) measure and 33.9% for the average bid-ask spread. When substituting the monthly bid-ask spread measure for large trades with the one calculated from small trades, the relation is still highly significant but with moderately decreased R^2_{adj} of 18.3% for the Schultz (2001) and 25.1% for the average bid-ask spread measure. Most interestingly, if we include transaction costs from both small and large trades, the R^2_{adj} surpasses notably the R^2_{adj} of the specifications in which only the institutional trades are employed. These results show not only that small trades are valuable for assessing institutional transaction costs but rather that the combination of both size categories offers superior information.

Because retail transaction costs are relevant for a large group of corporate bond investors, we also plug in daily retail-sized transaction costs $tc^{\text{small}}_{i,d,t}$ as the left-hand side of Equation (A.1). The results in Panel C of Table A.1 show that again both size categories carry a significant information content. As in Panel B, the R^2_{adj} are higher when explaining bid-ask spreads using other similarly sized trades. Finally, we again observe the highest R^2_{adj} when employing monthly spreads of both size categories. Summarizing, incorporating past transactions of small and large trades provides the most accurate assessment of a bond's liquidity.

A.3 Adjustments for the Average Bid-Ask Spread

The iterative two-stage weighted regression introduced in Section 2.2.2 relies on observations entering on a per trade basis into both stages. This means that when estimating the market-wide cost function, one transaction cost observation is uniquely assigned to one volume. And for the second step, when estimating the individual scaling factor, each trade is weighted with its volume category weight. In case of the average bid-ask spread measure, such an assignment is not possible as all trades in a bond i on a day d are combined to a single observation of $AvgBidAsk_{i,d}$. Therefore, we have to make two adjustments to

the iterative estimation procedure.

First, when estimating the market-wide transaction cost function, we face the problem that the nonparametric regression model only allows to estimate functions of the general form $y = \hat{m}(x)$. In other words, each bid-ask spread (on the left-hand side) has to be uniquely assigned to a trading volume vol. Since the function $c(vol)$ is not linear in vol, we cannot assign the average trade size of day d to $AvgBidAsk_{i,d}$. Therefore, we calculate $AvgBidAsk_{i,d}$ in Equation (2.4) for each traded volume on the day separately and assign to this observation the corresponding trade size. As this adjustment can lead to several observations representing the same day, we correct the weights of these observations such that each day contributes equally to the estimation of the market-wide transaction cost function. To do so, the weight of each observation is obtained by the product of its volume category weight (see Section 2.2.2) and the ratio of how often this volume v is traded compared to all traded volumes on the respective day for which we can calculate the size-dependent bid-ask spread. Again, to match the imbalance between sell and buy trades on a trading day, we calculate this ratio separately for both sides. The final ratio is then given by $\frac{1}{2}\left(\frac{(\#\ trades\ with\ volume\ v)_{i,d}^{buy}}{(\#\ trades)_{i,d}^{buy}} + \frac{(\#\ trades\ with\ volume\ v)_{i,d}^{sell}}{(\#\ trades)_{i,d}^{sell}}\right)$. Note that calculating the daily average bid-ask spread for each volume separately requires at least one sell and one buy trade with the same volume. Thus, the sample for the estimation of the market-wide transaction cost function in the first step of the iteration is only a subset of the sample used for the estimation of the individual scaling factors in the second step. However, since the estimation of the market-wide transaction cost function combines data from all bonds, the subset remains large enough to ensure a reliable estimation.

Second, since the bid-ask spread $AvgBidAsk_{i,d}$ on one day represents several volume categories, we have to adjust the weighting in the second step of the iteration. Given the weights of the volume categories $w(\cdot)$, we calculate the observation-day weight as the average of these weights, i.e., $\frac{1}{2}\left(\frac{1}{n_{i,d}^{buy}}\sum_{k=1}^{n_{i,d}^{buy}} w(vol_{k,i,d}^{buy}) + \frac{1}{n_{i,d}^{sell}}\sum_{k=1}^{n_{i,d}^{sell}} w(vol_{k,i,d}^{sell})\right)$.

For the estimation with a parametric functional form in Section 2.5.2, the first step is not necessary as Equation (2.10) is linear in the coefficients c_i. Instead, in the spirit of Edwards, Harris, and Piwowar (2007), we multiply for both measures the volume category weight with the inverse of this category's average squared residual and use this new weight in the first step of each iteration. This correction is necessary as the estimation noise is much larger for retail-sized trades. Without the correction, retail-sized trades would thus have a much larger influence on the OLS estimation of the market-wide function, leading to a relatively bad fit for large trades. For the nonparametric approach, such a correction

is not necessary as the nonparametric estimation procedure automatically decreases the weights for volumes farther away from the point that is estimated.

A.4 Bootstrapping Methodology

In our bootstrap-like procedure of Section 2.3.1, we use a subset of the trades from highly traded bonds to find out how liquidity measures can accommodate situations when bonds trade only scarcely or moderately. We select the trades that are part of the subset so that they resemble the trading pattern of scarcely traded bonds. We then compare the liquidity measure based on the subset of trades to the benchmark liquidity calculated from all available trades.

We define highly traded bonds as the 1% of bond-month observations with the highest number of trades. To ensure that liquidity measures do not suffer initially from a one-sided trading pattern, we require that the highly traded bond-month observations have at least one trade in each of the volume segments of Section 2.2.2. This leaves us with $H^{Schultz} = 3,880$ and $H^{AvgBidAsk} = 5,230$ highly-traded bond-month observations in case of the Schultz (2001) and the average bid-ask spread measure, respectively.

For each trading activity category and for both standard and size-adapted liquidity measures, the bootstrapping consists of four steps:

1. Randomly assign each highly-traded observation a bond-month observation of a scarcely traded bond that falls into the trading activity category. This step yields H pairs consisting each of a bond-month observation with a large number of trades and a bond-month observation with a small number of trades.

2. Select for each pair the most appropriate trades of the highly-traded observation to mirror the trading pattern of the scarcely traded bond. The best matching trade is selected from the trades with the same trade side in the same volume segment having the lowest squared difference of the trade sizes.

3. Calculate the liquidity measure based on both the full set of trades (i.e., the benchmark) and the mirrored trading pattern. To mitigate the impact of outliers, we winsorize both measures using the 1% and 99% quantile of the full cross-section of bonds from the benchmark's observation month.

4. Calculate the root-mean-squared percentage measurement error (RMSPE) as the difference between the two estimates across the H pairs.

We repeat these steps $n = 100$ times and obtain the final results as the average RMSPE across the 100 runs.

A.5 Individual Beta Estimation

In Section 2.4.3, we perform triple sorts, sorting on credit risk, liquidity, and liquidity betas. To estimate a bond's sensitivity to the corporate bond market liquidity risk, we use a Vasicek (1973)-like Bayesian approach.[74] In this approach, a bond's liquidity beta is estimated as the variance-weighted average of the individual beta and the prior, i.e., the beta of a double-sorted portfolio the bond is assigned to. The individual beta is estimated (if possible) by regressing the bond's returns on innovations in corporate bond market liquidity.

To ensure that not all bonds from the same double-sorted portfolio have the same prior beta, we need a finer grid for this (auxiliary) double sort. Thus, we sort bonds in the first stage into credit rating quintiles and in the second stage into liquidity quintiles based on a bond's amount outstanding. For the overlapping triple sort of the second approach, we follow Bongaerts, de Jong, and Driessen (2017) and use seven categories for the ratings (AAA, AA, A, BBB, BB, B, CCC) and PD quintiles. The classification of liquid and illiquid bonds is identical to the second step in the triple-sorting procedure.[75] A bond is again assigned to several portfolios simultaneously. Thus, we finally calculate the prior beta as the average of the betas from all portfolios the bond is assigned to.

Given the returns of the auxiliary portfolios and (if possible) the individual bond returns, we estimate the portfolio and individual betas using a rolling window of 24 months for which we require at least 12 observations.

[74] For details, see Appendix B of Bongaerts, de Jong, and Driessen (2017).

[75] Note that we do not sort bonds with AAA and CCC rating by their liquidity, as the number of bonds in both portfolios is not sufficient for a further classification.

Table A.2: **Fama-MacBeth analysis: Descriptives**
This table reports cross-sectional descriptive statistics of the panel data for the cross-sectional regression (2.8). For all variables, we first calculate the average across time, the statistics are then computed from the cross-section of portfolios. Sensitivities are estimated for the risk factors equity market return (EQ), shocks in equity market liquidity (EQLIQ), and shocks in corporate bond market liquidity (CBLIQ). Cross-sectional statistics are calculated based on the non-overlapping portfolio sort on credit quality, amount outstanding, and liquidity beta (described in detail in Section 2.4.3). In Panels A and B, we use the standard and size-adapted Schultz (2001) or average bid-ask-spread measure to calculate bond transaction costs and liquidity betas.

Panel A: Cross-sectional descriptives – Schultz (2001)

	Standard measure					Size-adapted measure				
	Mean	Std. dev.	$Q_{5\%}$	$Q_{50\%}$	$Q_{95\%}$	Mean	Std. dev.	$Q_{5\%}$	$Q_{50\%}$	$Q_{95\%}$
$E[r_{i,t}]$ (%)	2.26	1.28	0.81	1.87	5.04	2.26	1.27	0.88	1.90	5.03
β^{EQ}	0.1549	0.1017	0.0341	0.1511	0.3846	0.1550	0.0968	0.0399	0.1465	0.3885
β^{EQLIQ}	-0.0922	0.0544	-0.1702	-0.0854	-0.0143	-0.0730	0.0449	-0.1508	-0.0761	0.0090
β^{CBLIQ}	-0.1429	0.0532	-0.2516	-0.1376	-0.0650	-0.1841	0.0656	-0.2935	-0.1946	-0.0921
c (%)	1.34	0.37	0.70	1.29	1.98	1.11	0.27	0.62	1.12	1.55

Panel B: Cross-sectional descriptives – Average bid-ask spread

	Standard measure					Size-adapted measure				
	Mean	Std. dev.	$Q_{5\%}$	$Q_{50\%}$	$Q_{95\%}$	Mean	Std. dev.	$Q_{5\%}$	$Q_{50\%}$	$Q_{95\%}$
$E[r_{i,t}]$ (%)	2.26	1.28	0.87	1.89	5.15	2.26	1.29	0.84	1.87	5.12
β^{EQ}	0.1494	0.0972	0.0451	0.1308	0.3880	0.1546	0.1015	0.0458	0.1234	0.4090
β^{EQLIQ}	-0.0895	0.0474	-0.1491	-0.0995	-0.0131	-0.0692	0.0382	-0.1339	-0.0720	-0.0010
β^{CBLIQ}	-0.1352	0.0575	-0.2558	-0.1224	-0.0593	-0.1708	0.0646	-0.2772	-0.1590	-0.0806
c (%)	1.32	0.37	0.80	1.27	2.06	1.10	0.26	0.65	1.15	1.51

A.6 Fama-MacBeth Descriptive Statistics

Table A.2 reports cross-sectional summary statistics for the panel of expected returns, betas, and transaction costs of the second step cross-sectional regression (2.8) for the non-overlapping triple sort (the first approach in Section 2.4.3, which is our main approach).[76] Panel A presents the statistics for the standard and size-adapted Schultz (2001) measure and Panel B for the average bid-ask spread measures. The choice of the corporate bond market liquidity measure has only a slight impact on the distribution of the expected excess returns across the triple-sorted portfolios. In the same way, portfolios show a similar distribution regarding the size of transaction costs for both Schultz (2001) and average bid-ask spread measures. For the equity market beta β^{EQ}, we find positive sensitivities with a mean beta of roughly 0.15 and a standard deviation of 0.10, indicating that corporate bond returns decline in equity market downturns.

Regarding the market liquidity risk sensitivities, we find for almost all beta quantiles strictly negative estimates, indicating that corporate bond returns drop when equity or bond market illiquidity rises. The patterns, however, differ for the standard and the size-adapted versions. In case of the standard measures, the mean equity market liquidity beta β^{EQLIQ} and the corporate bond market liquidity beta β^{CBLIQ} are roughly -0.09 and -0.14, respectively. In case of the size-adapted measures, the absolute mean β^{EQLIQ} drops by more than 20% and the absolute mean β^{CBLIQ} increases by more than 25%. The (absolutely) larger equity market liquidity betas β^{EQLIQ} for the non-size adapted measures show that because of their correlation, an inaccurate measurement of bond market liquidity can increase the loadings of equity market liquidity.

Table A.3 reports cross-sectional correlations of expected excess returns, betas, and transaction costs. In Panel A, we depict the results for the Schultz (2001) measures and in Panel B for the average bid-ask spread measures. Independent of the choice of the corporate bond liquidity measure, we find a comparably low negative correlation (-0.15 to -0.18) between the equity market and the equity market liquidity beta. Corporate bond market liquidity beta and liquidity level are mildly more correlated (-0.35 to -0.53). Expected excess returns are correlated with equity market and equity market liquidity betas. Corporate bond liquidity also plays an important role for expected excess returns. Interestingly, all correlations of expected excess returns with bond market liquidity risk

[76] Descriptive statistics and correlations for the other two portfolio settings are qualitatively and quantitatively very similar to the ones in Tables A.2 and A.3. For the model estimated on individual bonds, beta standard deviations are, as expected, strongly inflated compared to the portfolio approaches.

Table A.3: **Fama-MacBeth analysis: Average cross-sectional correlations**
This table reports average cross-sectional correlations of the variables in the cross-sectional regression (2.8). We compute pairwise cross-sectional correlations across portfolios for each month and report their average across time. The set of risk factors consists of the equity return (EQ), shocks in equity market liquidity (EQLIQ), and shocks in corporate bond market liquidity (CBLIQ). Cross-sectional correlations are calculated based on the non-overlapping portfolio triple sort on credit quality, amount outstanding, and liquidity beta (described in detail in Section 2.4.3). In Panels A and B, we use the standard and size-adapted Schultz (2001) or average bid-ask-spread measure to calculate bond transaction costs and liquidity betas.

Panel A: Cross-sectional correlation – Schultz (2001)

	Standard measure					Size-adapted measure				
	$E[r_{i,t+1}]$	β^{EQ}	β^{EQLIQ}	β^{CBLIQ}	c	$E[r_{i,t+1}]$	β^{EQ}	β^{EQLIQ}	β^{CBLIQ}	c
$E[r_{i,t+1}]$	1	0.71	-0.30	-0.46	0.54	1	0.70	-0.23	-0.54	0.67
β^{EQ}		1	-0.18	-0.34	0.42		1	-0.18	-0.42	0.57
β^{EQLIQ}			1	0.16	-0.28			1	0.02	-0.28
β^{CBLIQ}				1	-0.35				1	-0.40
c					1					1

Panel B: Cross-sectional correlation – Average bid-ask spread

	Standard measure					Size-adapted measure				
	$E[r_{i,t+1}]$	β^{EQ}	β^{EQLIQ}	β^{CBLIQ}	c	$E[r_{i,t+1}]$	β^{EQ}	β^{EQLIQ}	β^{CBLIQ}	c
$E[r_{i,t+1}]$	1	0.71	-0.28	-0.54	0.47	1	0.73	-0.23	-0.62	0.65
β^{EQ}		1	-0.17	-0.36	0.34		1	-0.15	-0.57	0.59
β^{EQLIQ}			1	0.21	-0.24			1	0.10	-0.20
β^{CBLIQ}				1	-0.40				1	-0.53
c					1					1

β^{CBLIQ} and liquidity level c are (absolutely) stronger when using our more precise size-adapted liquidity measures.

A.7 Adapting the Repeat-Sales Measure

Adapting the repeat-sales measure of Bongaerts, de Jong, and Driessen (2017) requires minor changes to the iterative two-stage procedure of Sections 2.2.2 and 2.2.3. However, employing the basic measurement approach follows the same intuition.

Starting with the unadapted measure, estimating a trade's transaction costs is based on the idea that trade prices consist of a bond's fundamental value adjusted for transaction costs (see Section 1.3 of Bongaerts, de Jong, and Driessen, 2017). As a result, the log-difference between two consecutive prices $P_{k-1,i}$ and $P_{k,i}$ for bond i in portfolio j is the change in the fundamental value and the transaction costs , i.e.,

$$\ln(P_{k,i}) - \ln(P_{k-1,i}) = \sum_{s=t_{k-1}+1}^{t_k} R_{j,s} + c_j \cdot (Q_{k,i} - Q_{k-1,i}) + e_{k,i}. \tag{A.2}$$

The fundamental return in this equation is expressed by hourly latent portfolio returns $R_{j,s}$ between the two trade times t_{k-1} and t_k and an idiosyncratic term $e_{k,i}$. Transaction costs c_j in month t are assumed to be constant within a portfolio and are multiplied with each trade's direction $Q_{k,i}$, which equals 1 in case of a buyer-initiated trade and -1 for a seller-initiated trade.

Applying our basic approach in (2.1) to Equation (A.2), we replace the constant portfolio transaction costs with the market-wide cost function evaluated at the transaction's volume $c(vol_{k,i})$ multiplied by portfolio j's individual scaling factor $sf_j^{RepeatSales}$, leading to

$$\ln(P_{k,i}) - \ln(P_{k-1,i}) = \sum_{s=t_{k-1}+1}^{t_k} R_{j,s} + sf_j^{RepeatSales} \cdot (c(vol_{k,i}) \cdot Q_{k,i} - c(vol_{k-1,i}) \cdot Q_{k-1,i}) + e_{k,i}. \tag{A.3}$$

We estimate the scaling factor $sf_j^{RepeatSales}$ in (A.3) using the iterative two-stage procedure of Sections 2.2.2 and 2.2.3.[77] To prevent concerns of a look-ahead bias in Section 2.4, we

[77] In the first iteration, we perform an additional step and estimate a prior for the hourly portfolio returns using the standard measure. Thus, we can estimate the transaction cost function by using the return priors and by setting all scaling factors to 1.

estimate in the first step the transaction cost function $c(\cdot)$ using only the observations of the first quarter (see also footnote 22). In the second step, we follow Bongaerts, de Jong, and Driessen (2017) and estimate portfolio j's hourly returns and monthly scaling factors for an entire quarter in a pooled regression. Consistent with the two other size-adapted measures, the adapted repeat-sales measure can be estimated whenever the data requirements of the standard measure are satisfied.

Similar to the average bid-ask-spread measure (see Appendix A.3), the nonparametric estimation of the transaction cost function leads to the problem that trades do not enter separately into Equation (A.3) but as two consecutive trades. Hence, to estimate the market-wide transaction cost function in the first step of the iteration, we can only employ log-differences between two consecutive trades if they share the same volume. Despite the resulting sample reduction, using data of all bonds still ensures a reliable estimation of the market-wide transaction cost function.

Appendix B

Additional Information on *Expected Bond Liquidity*

B.1 Bond Data Filters and Yield Spread Calculation

We use transaction data from Enhanced TRACE for the U.S. corporate bond market for the time period from October 1, 2004 to June 30, 2017. Despite TRACE being gradually introduced starting in 2002 by the FINRA, we restrict our examination period to a start in October 2004. After this date, market participants had to report almost all trades. To clean the data, we apply several filters comparable to the ones used in Schestag, Schuster, and Uhrig-Homburg (2016) and in Chapter 2. In a first step, we apply the standard procedures of Dick-Nielsen (2009, 2014) to remove duplicates, withdrawn and corrected entries. In the second step, we apply the median and reversal filter of Edwards, Harris, and Piwowar (2007) to exclude erroneous entries and extreme outliers. Finally, we demand bonds to satisfy several conditions. First, we exclude bonds with special features (perpetuals, convertible and puttable bonds, floating coupon payments). Second, bonds have to be senior unsecured, USD denominated, and there shall be no entity backing the bond with a guarantee. Third, we demand bonds to be actively traded in at least twelve months within our sample period or in 50% of the months they are active. Last, we exclude observations of defaulted bonds after the default happened. Overall, our sample consists of 25,918 bonds and 61,360,046 trades, which is approximately 57% of the about 108 million trades during our observation period.

We require a bond's yield spread for the analysis in Section 3.2.5. If the yield-to-

maturity is larger than the yield-to-call, the reported yield in TRACE often (but not always) corresponds to the yield-to-call. We calculate both yields and select the one that is closest to the reported yield. We drop observations for which both differences are larger than 1 basis point (about 0.7% of all observations). We discount the bond's cash flows with the risk-free Treasury curve to calculate the price of an artificial Treasury bond with the same cash flow structure (using updated data from Gürkaynak, Sack, and Wright (2007) published by the Federal Reserve on http://www.federalreserve.gov). Finally, the yield spread is defined as the continuously compounded yield computed from the reported price minus the corresponding yield calculated from the price of the artificial Treasury bond (see also Gehde-Trapp, Schuster, and Uhrig-Homburg, 2018).

B.2 Mutual Fund Data

We use mutual fund data from Morningstar. We identify corporate bond funds using the Morningstar category classifications and the fund's prospectus objective. Specifically, funds have to be in one of the categories 'US Fund Long-Term Bond', 'US Fund Intermediate Core Bond', 'US Fund Intermediate Core-Plus Bond', 'US Fund Short-Term Bond', 'US Fund Ultrashort Bond', 'US Fund Corporate Bond', 'US Fund High Yield Bond', 'US Fund Target Maturity', or 'US Fund Multisector Bond'. Additionally, we exclude funds with a focus on government or municipal bonds in their prospectus objective. The remaining filters follow Goldstein, Jiang, and Ng (2017). We exclude index funds as well as fund share classes during their first year. Finally, to get sufficient data coverage when merging fund holding data with TRACE, we restrict the sample in our analyses to the time period from January 2008 to June 2017. After applying these filters, our sample consists of 3,492 share classes from 1,005 corporate bond funds.

Appendix C

Additional Information on *Comparing Forecast Performance of Finance Panel Data Models*

C.1 Pre-clustering with hierarchical clustering

We identify in Section 4.2 rectangular clusters in cross-section and time by employing the widely used hierarchical clustering algorithm (see, e.g., Panton, Lessig, and Joy, 1976; Bouvatier, Lepetit, and Strobel, 2014). The main idea of this bottom-up clustering algorithm is to iteratively merge clusters of observations with the lowest distance to each other. Starting with each observation as a separate cluster, the iteration procedure stops once all observations are assigned to the same cluster. The result is a tree called dendogram where each cluster agglomeration is represented by a node (with the starting clusters as terminal nodes and the final cluster containing all observations as root) and the tree's branches indicate the merging distances. In the next step, the tree is cut at a desired height. The tree nodes which are below this height then form the final cluster of the clustering algorithm.

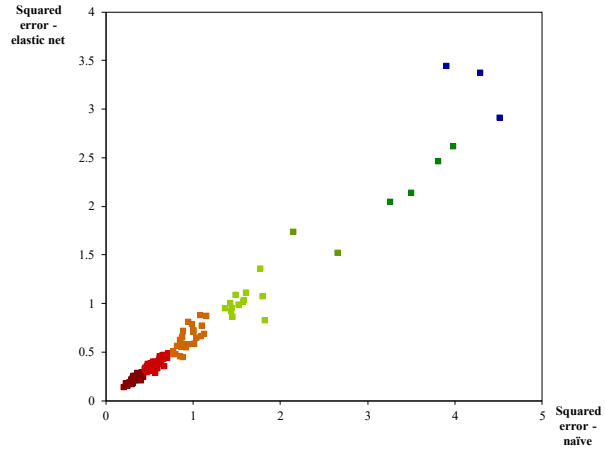
In particular, we detect empirically the rectangular clusters as intersection of separately identified time and cross-sectional clusters. To this end, we obtain time clusters by employing the hierarchical clustering to the time series of average cross-sectional prediction errors of two competing models and respectively the bond clusters based on the cross-section of average prediction errors across time. Merging observations or clusters in each iteration to a larger cluster follows Ward's method. Finally, we get the final cluster

allocation by cutting the dendogram at a set percentage of the maximum height where the specific cutoff value depends on the empirical application.

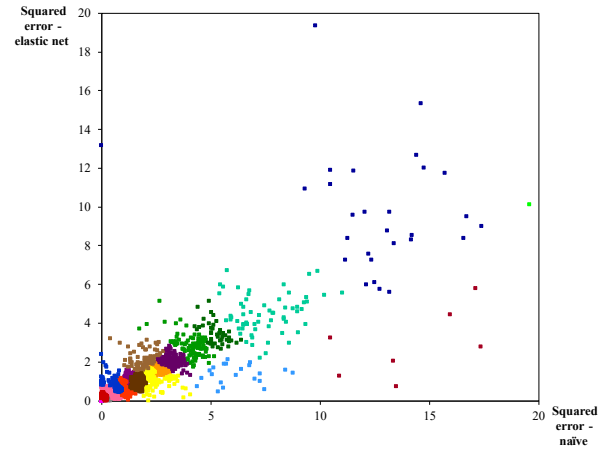
For the sake of brevity, we restrict our clustering results of the hierarchical clustering in Section 4.3.2 to the pairwise comparisons between the dynamic model employing stepwise regression and either the naïve approach or the dynamic model based on the elastic net procedure. The hierarchical clustering results for the remaining model pairs are presented in Figure C.1. While Panels A and B show the detected clusters of the naïve forecast errors and those of the dynamic models based on the elastic net procedure and in-sample significance, Panel C shows the clusters when comparing the dynamic stepwise and in-sample significance models. Lastly, Panel D shows the pre-clustering of the forecast errors for the elastic net and the in-sample significance model. Consistent with the previous results of Section 4.3.2, we find that the forecast errors in the time series of the naïve and the dynamic models exhibit an almost linear behavior, while the distribution in the cross-sectional clusters is much more spread. Again, when pairwise clustering the forecast errors of the remaining dynamic models, we find that the clustering algorithm is capable of identifying the respective heterogeneity in the cross-sectional or time dimension and assigns observations to separate clusters accordingly.

Panel A: Pre-clustering naïve vs. elastic net

Time cluster

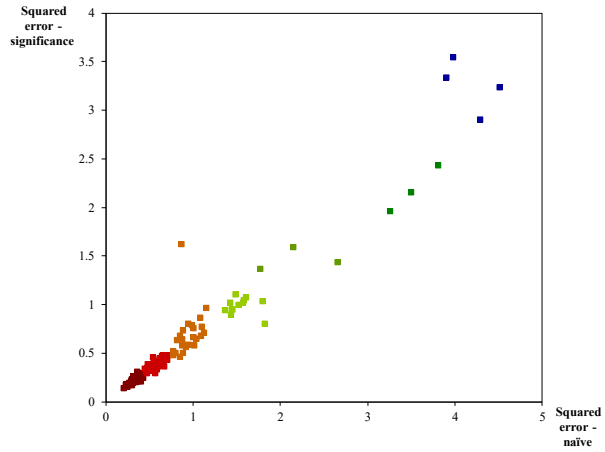


Bond cluster

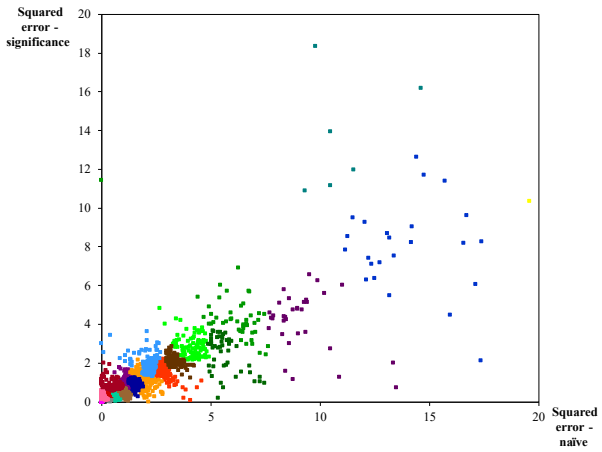


Panel B: Pre-clustering naïve vs. in-sample significance

Time cluster

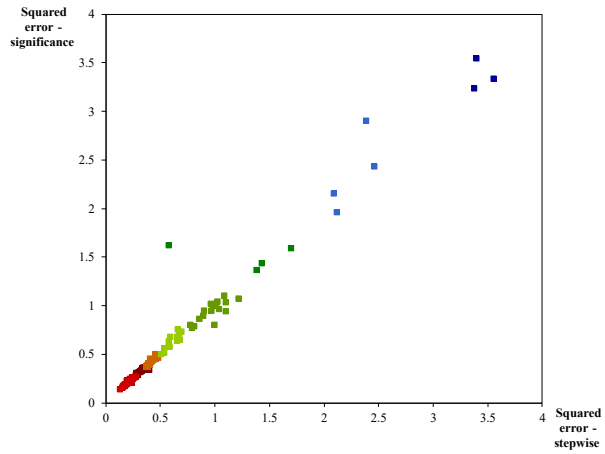


Bond cluster

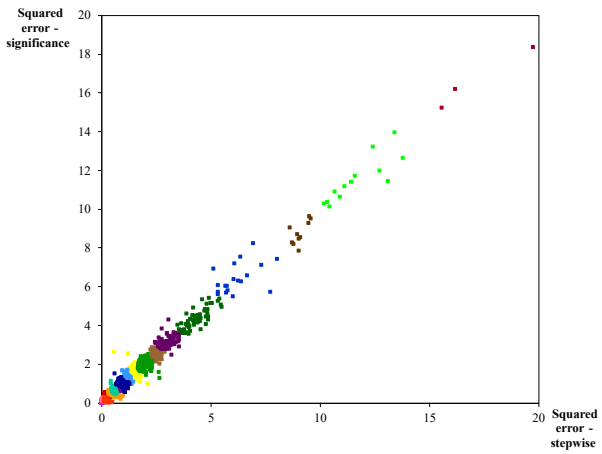


Panel C: Pre-clustering stepwise vs. in-sample significance

Time cluster



Bond cluster



Panel D: Pre-clustering elastic net vs. in-sample significance

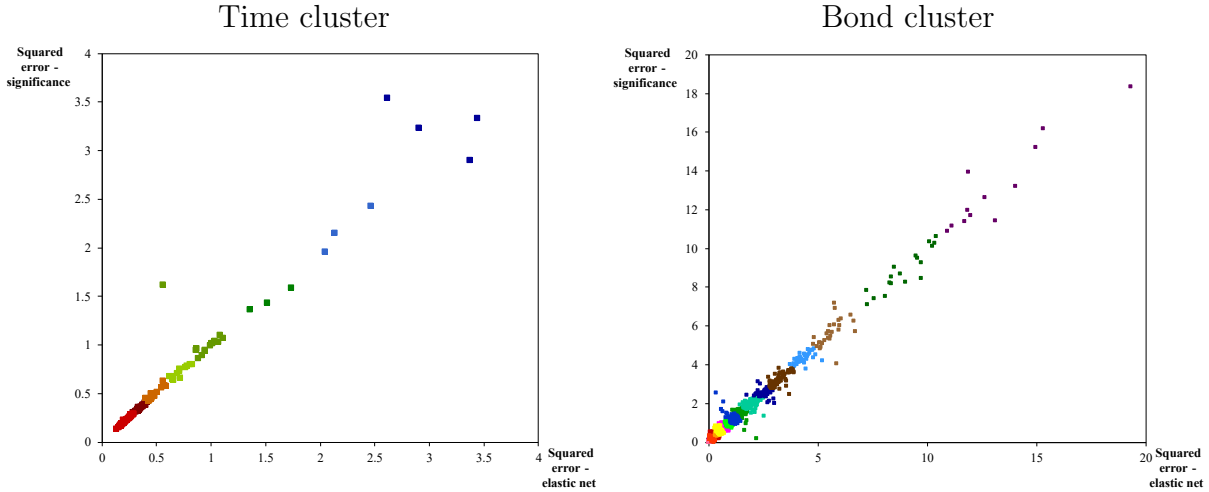


Figure C.1: **Remaining pre-clustering across time and bonds**
This figure shows the pre-clustering across time (left image) and bonds (right image) for the remaining comparisons of Section C.1 for the individual bond liquidity forecast models of Section 4.3.1. Time and cross-sectional clusters are identified by either employing hierarchical clustering on the time series or the of average cross-sectional squared prediction errors of the competing models or on the cross-section of the bonds' mean squared prediction error across time. In Panel A and B, bond liquidity is once predicted by the naïve approach and second either by the dynamic forecast model based on the elastic net or on in-sample significance. In Panel C and D, bond liquidity is first predicted by the forecast model based on in-sample significance and second by one of the holdout procedure models, stepwise regression or elastic net.

Appendix D

Out-of-Sample Yield Spread Regression Tests with Pre-Clustering

In Section 2.3.1, we calculate implied yield spread changes in an out-of-sample setting and compare the mean squared error (MSE) with respect to the actual yield spread change between the models based on (2.6) when either employing a standard liquidity measure or its size-adapted counterpart. We perform a similar exercise in Section 3.2.5 but focus on the difference in the MSE when employing either current liquidity or forecasts of our prediction approach of Section 3.2 as proxy for expected liquidity in the regression model (3.3). In both exercises, we test the differences in the MSE using a Diebold and Mariano (1995) test in the spirit of Harvey, Leybourne, and Newbold (1997). However, this test is similar to the first alternative of the overall Diebold and Mariano (1995) test in Section 4.3.3 and thus does not account for potential two-dimensional heterogeneity in the out-of-sample errors in the cross-section and the time space. In order to rule out that our test decisions are driven by inflated t-statistics, we repeat the out-of-sample tests and apply our clustered Diebold and Mariano (1995) test of Section 4.2.2 to test for equal accuracy. Consistent with our in-sample standard errors, we pre-specify each pair of observation month and bond as a separate rectangular cluster.[78]

The test results for the analysis of Section 2.3.1 are presented in Panel A of Table D.1. Consistent with the results in Table 2.3, we find that the MSE drops significantly once we add a standard or size-adapted liquidity measure to the baseline regression that only

[78]Note that this is approach leads to clusters containing only one observation. However, if we are more conservative and detect clusters data-driven by employing hierarchical clustering, our results are qualitatively comparable.

Table D.1: **Yield spread regression tests with pre-clustering**

This table reports test results for the pairwise comparisons of the out-of-sample mean squared errors (MSE) of the panel regressions of Sections 2.3.1 and 3.2.5 when we employ the pooled hypothesis test with pre-clustered standard errors of Chapter 4. In Panel A, we compare the MSE when we employ either standard liquidity measures or their size-adapted counterparts. In particular, we compare a specification to the baseline specification without a liquidity measure and in case of the size-adapted measure additionally to the specification of its standard measure. In Panel B, we compare the MSE when employing either current liquidity or expected liquidity. Expected liquidity is either estimated via the linear combination model of Section 3.2.3 or via the random forest model of Section 3.4.1 and we measure bond liquidity using the average bid-ask spread measure. We first compare a specification to the baseline specification without a liquidity measure and in case of an expected liquidity proxy additionally to the specification employing current liquidity. Each month and bond pair marks a separate rectangular cluster in the pre-step. The t-statistics are given in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% level.

Panel A: Size-adapted liquidity measure

	Schultz (2001)			Average bid-ask spread		
	Baseline	Standard	Size-adapted	Baseline	Standard	Size-adapted
MSE	0.909	0.905	0.902	1.683	1.670	1.656
Δ(MSE)		-0.004***	-0.007***/-0.003*		-0.013**	-0.027***/-0.014***
		(3.11)	(2.59)/(1.87)		(2.34)	(2.87)/(2.94)

Panel B: Expected liquidity

	Baseline	Current liquidity	Exp. liquidity - lin. combination	Exp. liquidity - random forest
MSE	0.942	0.939	0.918	0.915
Δ(MSE)		-0.003	-0.024/-0.021	-0.027**/-0.024**
		(1.47)	(1.61)/(1.63)	(2.08)/(2.08)

includes lagged yield spread changes and control variables. Comparing the specifications that employ either a standard liquidity measure or its size-adapted counterpart, we find that the additional drop in the MSE caused by the size-adaption is significant at the 10% level in case of the Schultz (2001) measure and at the 1% level in case of the average bid-ask spread measure. Lastly, the results for the analysis of Section 3.2.5 are presented in Panel B of Table D.1. Interestingly, we find that the MSE of the baseline regression now is not significantly lower compared to the specification that additionally includes changes in the current liquidity as naïve proxy for expected liquidity. When incorporating expected liquidity estimates based on the linear combination model of Section 3.2.3 the drop in the MSE is surprisingly slightly insignificant too (p-value of 10.7%) as well as the additional drop in the MSE compared to the current liquidity specification (p-value of 10.4%). However, if we estimate expected liquidity with the random forest model of Section 3.4.1 that offers the overall lowest MSE, we find that this specification is associated with a significantly higher precision compared to both, the baseline regression and the specification employing current liquidity as proxy for expected liquidity. Thus, the clustered Diebold and Mariano (1995) test results still support the finding that the forecasts of our prediction approach are able to better explain yield spread changes than the literature's naïve proxy.

Bibliography

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge, 2017, When Should You Adjust Standard Errors for Clustering?, Working paper.

Acharya, V. and L. H. Pedersen, 2005, Asset Pricing with Liquidity Risk, *Journal of Financial Economics* 77, 375–410.

Acharya, V. V., Y. Amihud, and S. T. Bharath, 2013, Liquidity Risk of Corporate Bond Returns: Conditional Approach, *Journal of Financial Economics* 110, 358–386.

Adrian, T., M. Fleming, O. Shachar, and E. Vogt, 2017, Market Liquidity after the Financial Crisis, *Annual Review of Financial Economics* 9, 43–83.

Agarwal, V. and N. Naik, 2004, Risks and Portfolio Decisions Involving Hedge Funds, *The Review of Financial Studies* 17, 63–98.

Ambrose, B. W., K. N. Cai, and J. Helwege, 2008, Forced Selling of Fallen Angels, *The Journal of Fixed Income* 18, 72–85.

Amihud, Y., 2002, Illiquidity and Stock Returns: Cross-Section and Time Series Effects, *Journal of Financial Markets* 5, 31–56.

Amihud, Y. and H. Mendelson, 1986, Asset Pricing and the Bid-Ask Spread, *Journal of Financial Economics* 17, 223–249.

Amihud, Y. and H. Mendelson, 1991, Liquidity, Maturity, and the Yields on U.S. Treasury Securities, *Journal of Finance* 46, 1411–1425.

Anderson, M. and R. M. Stulz, 2017, Is Post-Crisis Bond Liquidity Lower?, Working paper.

Ang, A., J. Liu, and K. Schwarz, 2020, Using Stocks or Portfolios in Tests of Factor Models, *Journal of Financial and Quantitative Analysis* 55, 709–750.

Athey, S., J. Tibshirani, and S. Wager, 2019, Generalized Random Forests, *The Annals of Statistics* 47, 1148–1178.

Bai, J., T. G. Bali, and Q. Wen, 2019, Common Risk Factors in the Cross-Section of Corporate Bond Returns, *Journal of Financial Economics* 131, 619–642.

Bai, J. and P. Collin-Dufresne, 2011, The Determinants of the CDS-Bond Basis during the Financial Crisis of 2007-2009, Working paper.

Bao, J., M. O'Hara, and X. A. Zhou, 2018, The Volcker Rule and Corporate Bond Market Making in Times of Stress, *Journal of Financial Economics* 130, 95–113.

Bao, J., J. Pan, and J. Wang, 2011, The Illiquidity of Corporate Bonds, *Journal of Finance* 66, 911–946.

Behrens, C., C. Pierdzioch, and M. Risse, 2018, A Test of the Joint Efficiency of Macroeconomic Forecasts using Multivariate Random Forests, *Journal of Forecasting* 37, 560–572.

Ben-Rephael, A., O. Kadan, and A. Wohl, 2015, The Diminishing Liquidity Premium, *Journal of Financial and Quantitative Analysis* 50, 197–229.

Berndt, A., 2015, A Credit Spread Puzzle for Reduced-Form Models, *Review of Asset Pricing Studies* 5, 48–91.

Bernoth, K. and A. Pick, 2011, Forecasting the Fragility of the Banking and Insurance Sectors, *Journal of Banking & Finance* 35, 807–818.

Bessembinder, H., S. Jacobsen, W. Maxwell, and K. Venkataraman, 2018, Capital Commitment and Illiquidity in Corporate Bonds, *Journal of Finance* 73, 1615–1661.

Bongaerts, D., F. de Jong, and J. Driessen, 2017, An Asset Pricing Approach to Liquidity Effects in Corporate Bond Markets, *Review of Financial Studies* 30, 1229–1269.

Bouvatier, V., L. Lepetit, and F. Strobel, 2014, Bank Income Smoothing, Ownership Concentration and the Regulatory Environment, *Journal of Banking & Finance* 41, 253–270.

Boyarchenko, Nina, Domenico Giannone, and Or Shachar, 2019, Flighty liquidity, Working paper.

Brogaard, J., D. Li, and Y. Xia, 2017, Stock Liquidity and Default Risk, *Journal of Financial Economics* 124, 486–502.

Brunnermeier, M. K. and L. H. Pedersen, 2009, Market Liquidity and Funding Liquidity, *Review of Financial Studies* 22, 2201–2238.

Capponi, Agostino, Paul Glasserman, and Marko Weber, 2020, Swing pricing for mutual funds: Breaking the feedback loop between fire sales and fund redemptions, *Management Science*.

Chakravarty, S. and A. Sarkar, 2003, Trading Costs in Three US Bond Markets, *The Journal of Fixed Income* 13, 39–48.

Chen, J., H. Hong, and J. C. Stein, 2001, Forecasting Crashes: Trading Volume, Past Returns, and Conditional Skewness in Stock Prices, *Journal of Financial Economics* 61, 345–381.

Chen, L., D. Lesmond, and J. Wei, 2007, Corporate Yield Spreads and Bond Liquidity, *Journal of Finance* 62, 119–149.

Chen, Q., I. Goldstein, and W. Jian, 2010, Payoff Complementarities, Self-fulfilling Beliefs and Mutual Fund Outflows, *Journal of Financial Economics* 98, 239–262.

Chernobai, A., P. Jorion, and F. Yu, 2011, The Determinants of Operational Risk in US Financial Institutions, *Journal of Financial and Quantitative Analysis* 46, 1683–1725.

Chinco, A., A. Clark-Joseph, and M.o Ye, 2019, Sparse Signals in the Cross-Section of Returns, *The Journal of Finance* 74, 449–492.

Choi, J. and Y. Huh, 2019, Customer Liquidity Provision: Implications for Corporate Bond Transaction Costs, Working paper.

Chordia, T., A. Goyal, Y. Nozawa, A. Subrahmanyam, and Q. 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.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, Commonality in liquidity, *Journal of financial economics* 56, 3–28.

Chordia, T., A. Sarkar, and A. Subrahmanyam, 2005, An Empirical Analysis of Stock and Bond Market Liquidity, *The Review of Financial Studies* 18, 85–129.

Chudik, A., G. Kapetanios, and M. H. Pesaran, 2018, A One Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models, *Econometrica* 86, 1479–1512.

Clark, T. E. and M. W. McCracken, 2001, Tests of Equal Forecast Accuracy and Encompassing for Nested Models, *Journal of Econometrics* 105, 85–110.

Corwin, S. and P. Schultz, 2012, A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices, *Journal of Finance* 67, 719–760.

Daniel, K., S. Titman, and K.C. Wei, 2001, Explaining the Cross-Section of Stock Returns in Japan: Factors or Characteristics?, *Journal of Finance* 56, 743–766.

Davis, J. L., E. F. Fama, and K. R. French, 2000, Characteristics, Covariances, and Average Returns: 1929 to 1997, *The Journal of Finance* 55, 389–406.

Dick-Nielsen, J., 2009, Liquidity Biases in TRACE, *Journal of Fixed Income* 19, 43–55.

Dick-Nielsen, J., 2014, How to Clean Enhanced TRACE Data, Working paper.

Dick-Nielsen, J., P. Feldhütter, and D. Lando, 2012, Corporate Bond Liquidity Before and After the Onset of the Subprime Crisis, *Journal of Financial Economics* 103, 471–492.

Dick-Nielsen, J. and M. Rossi, 2019, The Cost of Immediacy for Corporate Bonds, *Review of Financial Studies* 32, 1–41.

Diebold, F. and R. Mariano, 1995, Comparing Predictive Accuracy, *Journal of Business & Economic Statistics* 13, 253–263.

Downing, C., S. Underwood, and Y. Xing, 2005, Is liquidity Risk Priced in the Corporate Bond Market?, Working paper.

Duan, J., J. Sun, and T. Wang, 2012, Multiperiod Corporate Default Prediction – A Forward Intensity Approach, *Journal of Econometrics* 170, 191–209.

Duffee, G. R., 1998, The Relation between Treasury Yields and Corporate Bond Yield Spreads, *The Journal of Finance* 53, 2225–2241.

Ederington, L., W. Guan, and L. Yang, 2015, Bond Market Event Study Methods, *Journal of Banking & Finance* 58, 281–293.

Edwards, A. K., L. E. Harris, and M. S. Piwowar, 2007, Corporate Bond Market Transaction Costs and Transparency, *Journal of Finance* 62, 1421–1451.

Ellul, A., C. Jotikasthira, and C. T. Lundblad, 2011, Regulatory Pressure and Fire Sales in the Corporate Bond Market, *Journal of Financial Economics* 101, 596–620.

Fama, E. and J. MacBeth, 1973, Risk, Return, and Equilibrium: Empirical Tests, *Journal of political economy* 81, 607–636.

Feldhütter, P., 2012, The Same Bond at Different Prices: Identifying Search Frictions and Selling Pressure, *Review of Financial Studies* 25, 1155–1206.

Feldhütter, P., E. Hotchkiss, and K. Oğuzhan, 2016, The Value of Creditor Control in Corporate Bonds, *Journal of Financial Economics* 121, 1–27.

Feldhütter, P. and S. Schaefer, 2018, The Myth of the Credit Spread Puzzle, *The Review of Financial Studies* 31, 2897–2942.

Friewald, N., R. Jankowitsch, and M. G. Subrahmanyam, 2012, Illiquidity or Credit Deterioration: A Study of Liquidity in the US Corporate Bond Market During Financial Crisis, *Journal of Financial Economics* 105, 18–36.

Fuertes, A. and E. Kalotychou, 2007, Optimal Design of Early Warning Systems for Sovereign Debt Crises, *International Journal of Forecasting* 23, 85–100.

Gebhardt, W. R., S. Hvidkjaer, and B. Swaminathan, 2005, The Cross-Section of Expected Corporate Bond Returns: Betas or Characteristics?, *Journal of Financial Economics* 75, 85–114.

Gehde-Trapp, M., P. Schuster, and M. Uhrig-Homburg, 2018, The Term Structure of Bond Liquidity, *Journal of Financial and Quantitative Analysis* 53, 2161–2197.

Giacomini, R. and I. Komunjer, 2005, Evaluation and Combination of Conditional Quantile Forecasts, *Journal of Business & Economic Statistics* 23, 416–431.

Giacomini, R. and B. Rossi, 2010, Forecast Comparisons in Unstable Environments, *Journal of Applied Econometrics* 25, 595–620.

Giacomini, R. and H. White, 2006, Tests of Conditional Predictive Ability, *Econometrica* 74, 1545–1578.

Goldstein, I., H. Jiang, and D. T. Ng, 2017, Investor Flows and Fragility in Corporate Bond Funds, *Journal of Financial Economics* 126, 592–613.

Goyenko, R. and A. Ukhov, 2009, Stock and Bond Market Liquidity: A Long-Run Empirical Analysis, *Journal of Financial and Quantitative Analysis* 44, 189–212.

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka, 2009, Do Liquidity Measures Measure Liquidity?, *Journal of financial Economics* 92, 153–181.

Green, R. C., B. Hollifield, and N. Schürhoff, 2007a, Dealer Intermediation and Price Behavior in the Aftermarket for New Bond Issues, *Journal of Financial Economics* 86, 643–682.

Green, R. C., B. Hollifield, and N. Schürhoff, 2007b, Financial Intermediation and the Cost of Trading in an Opaque Market, *Review of Financial Studies* 20, 275–314.

Greenwich Associates, 2018, U.S. Treasurys Trade Electronically - But Where are the Algos?, `https://www.greenwich.com/blog/us-treasurys-trade-electronically-but-where-algos/`.

Greenwich Associates, 2019, Corporate Bond Trading in 2019, `https://www.greenwich.com/fixed-income/corporate-bond-trading-2019/`.

Gu, S., B. Kelly, and D. Xiu, 2020, Empirical Asset Pricing via Machine Learning, *The Review of Financial Studies* 33, 2223–2273.

Gürkaynak, R., B. Sack, and J. Wright, 2007, The US Treasury Yield Curve: 1961 to the Present, *Journal of Monetary Economics* 54, 2291–2304.

Harris, L. and M. Piwowar, 2006, Secondary Trading Costs in the Municipal Bond Market, *Journal of Finance* 61, 1361–1397.

Harvey, D., S. Leybourne, and P. Newbold, 1997, Testing the Equality of Prediction Mean Squared Errors, *International Journal of Forecasting* 13, 281–291.

Hasbrouck, J., 2009, Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data, *Journal of Finance* 64, 1445–1477.

He, Z. and K. Milbradt, 2014, Endogenous Liquidity and Defaultable Bonds, *Econometrica* 82, 1443–1508.

He, Z. and W. Xiong, 2012, Rollover Risk and Credit Risk, *The Journal of Finance* 67, 391–430.

Hendershott, T. and A. Madhavan, 2015, Click or Call? Auction versus Search in the Over-the-Counter Market, *The Journal of Finance* 70, 419–447.

Holm, S., 1979, A Simple Sequentially Rejective Multiple Test Procedure, *Scandinavian Journal of Statistics* 65–70.

Hong, G. and A. Warga, 2000, An Empirical Study of Bond Market Transactions, *Financial Analysts Journal* 56, 32–46.

Hotchkiss, E. and G. Jostova, 2017, Determinants of Corporate Bond Trading: A Comprehensive Analysis, *Quarterly Journal of Finance* 7, 1750003.

Huang, J., K. Wei, and H. Yan, 2007, Participation Costs and the Sensitivity of Fund Flows to Past Performance, *The Journal of Finance* 62, 1273–1311.

Jankowitsch, R., A. Nashikkar, and M. Subrahmanyam, 2011, Priced Dispersion in OTC Markets: A New Measure of Liquidity, *Journal of Banking and Finance* 35, 343–357.

Jin, Dunhong, Marcin T Kacperczyk, Bige Kahraman, and Felix Suntheim, 2019, Swing Pricing and Fragility in Open-End Mutual Funds, .

Kacperczyk, M., S. Nieuwerburgh, and L. Veldkamp, 2014, Time-varying Fund Manager Skill, *The Journal of Finance* 69, 1455–1484.

Kaushik, M. and B. Mathur, 2014, Comparative Study of K-Means and Hierarchical Clustering Techniques, *International Journal of Software and Hardware Research in Engineering* 2, 93–98.

Kozak, S., S. Nagel, and S. Santosh, 2020, Shrinking the Cross-Section, *Journal of Financial Economics* 135, 271–292.

Kyle, A. S., 1985, Continuous Auctions and Insider Trading, *Econometrica* 1315–1335.

Leuz, C., D. Nanda, and P. D. Wysocki, 2003, Earnings Management and Investor Protection: An International Comparison, *Journal of Financial Economics* 69, 505–527.

Lewellen, J., S. Nagel, and J. Shanken, 2010, A Skeptical Appraisal of Asset Pricing Tests, *Journal of Financial Economics* 96, 175–194.

Liang, K.-Y. and S. L. Zeger, 1986, Longitudinal Data Analysis using Generalized Linear Models, *Biometrika* 73, 13–22.

Lin, H., J. Wang, and C. Wu, 2011, Liquidity Risk and the Cross-Section of Expected Corporate Bond Returns, *Journal of Financial Economics* 99, 628–650.

Liu, L., H. R. Moon, and F. Schorfheide, 2020, Forecasting with Dynamic Panel Data Models, *Econometrica* 88, 171–201.

Liu, X. and J. R. Ritter, 2011, Local Underwriter Oligopolies and IPO Underpricing, *Journal of Financial Economics* 102, 579–601.

Loon, Y. C. and Z. K. Zhong, 2016, Does Dodd-Frank affect OTC Transaction Costs and Liquidity? Evidence from Real-Time CDS Trade Reports, *Journal of Financial Economics* 119, 645–672.

Lou, X. and T. Shu, 2017, Price Impact or Trading Volume: Why Is the Amihud (2002) Measure Priced?, *Review of Financial Studies* 30, 4481–4520.

Mahanti, S., A. Nashikkar, M. Subrahmanyam, G. Chacko, and G. Mallik, 2008, Latent Liquidity: A New Measure of Liquidity, with an Application to Corporate Bonds, *Journal of Financial Economics* 88, 272–298.

Newey, W. and K. West, 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–708.

NUS-RMI, 2016, NUS-RMI Credit Research Initiative Technical Report Version: 2016 Update 1, *Global Credit Review* 6, 49–132.

Panopoulou, E. and S. Vrontos, 2015, Hedge Fund Return Predictability; To Combine Forecasts or Combine Information?, *Journal of Banking & Finance* 56, 103–122.

Panton, D. B., V. P. Lessig, and O. M. Joy, 1976, Comovement of International Equity Markets: A Taxonomic Approach, *Journal of Financial and Quantitative Analysis* 415–432.

Pástor, L. and R. F. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy* 111, 642–685.

Patton, A. J., 2011, Volatility Forecast Comparison using Imperfect Volatility Proxies, *Journal of Econometrics* 160, 246 – 256.

Pesaran, M. and A. Timmermann, 1995, Predictability of Stock Returns: Robustness and Economic Significance, *The Journal of Finance* 50, 1201–1228.

Petersen, M. A., 2009, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, *The Review of Financial Studies* 22, 435–480.

Rapach, D., J. Strauss, and G. Zhou, 2010, Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy, *The Review of Financial Studies* 23, 821–862.

Reichenbacher, M. and M. Schienle, 2021, Comparing Forecast Performance for Finance Panel Data Models, Working paper.

Reichenbacher, M. and P. Schuster, 2020, Size-Adapted Bond Liquidity Measures and Their Asset Pricing Implications, Working paper.

Reichenbacher, M., P. Schuster, and M. Uhrig-Homburg, 2020, Expected Bond Liquidity, Working paper.

Romano, J. P. and M. Wolf, 2005, Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing, *Journal of the American Statistical Association* 100, 94–108.

Schestag, R., P. Schuster, and M. Uhrig-Homburg, 2016, Measuring Liquidity in Bond Markets, *Review of Financial Studies* 29, 1170–1219.

Schmidt, B., 2015, Costs and Benefits of Friendly Boards During Mergers and Acquisitions, *Journal of Financial Economics* 117, 424–447.

Schultz, P., 2001, Corporate Bond Trading Costs: A Peek Behind the Curtain, *Journal of Finance* 56, 677–698.

Schuster, P., 2020, Trade-Size Related Frictions in Bond Markets, Habilitation thesis, Karlsruhe Institute of Technology (KIT).

sifma, 2021a, US Equity Issuance and Trading Volumes, Securities Industry and Financial Markets Association, `https://www.sifma.org/resources/research/us-equity-stats/`.

sifma, 2021b, US Fixed Income Trading Volume, Securities Industry and Financial Markets Association, `https://www.sifma.org/resources/research/us-fixed-income-trading-volume/`.

Smith, J. and K. Wallis, 2009, A Simple Explanation of the Forecast Combination Puzzle, *Oxford Bulletin of Economics and Statistics* 71, 331–355.

Stoll, H. R., 1978, The Supply of Dealer Services in Securities Markets, *Journal of Finance* 33, 1133–1151.

Timmermann, A. and Y. Zhu, 2019, Comparing Forecasting Performance with Panel Data, Working paper.

Titman, S. and C.n Tiu, 2011, Do the Best Hedge Funds Hedge?, *The Review of Financial Studies* 24, 123–168.

Vasicek, O., 1973, A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas, *Journal of Finance* 28, 1233–1239.

Warga, Arthur, 1992, Bond returns, liquidity, and missing data, *Journal of Financial and Quantitative Analysis* 27, 605–617.

White, H., 1984, A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica* 48, 817–838.