Essays on Sustainable Finance and Regulatory Risk

Dissertation
zur Erlangung des akademischen Grades
Doctor rerum politicarum

am Fachbereich für Rechts- und Wirtschaftswissenschaften der
Technischen Universität Darmstadt

vorgelegt von
Lukas Müller

Erstgutachter: Prof. Dr. Dirk Schiereck
Zweitgutachter: Ass.-Prof. Dr. Sebastian Stöckl

Darmstadt, April 2024

Müller, Lukas: Essays on Sustainable Finance and Regulatory Risk
Darmstadt, Technische Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUprints: 2024
Tag der mündlichen Prüfung: 15.04.2024

Abstract

Natural experiments have a long tradition in empirical corporate finance. Event study methodology is an incumbent approach for measuring the economic impact of an event on shareholder wealth. Within this context, this dissertation covers two main subject areas. The first part offers a methodological contribution to fundamental research in the field of corporate bonds. In fixed-income research, event study methodology lags far behind event studies in the field of equity, and existing approaches are insufficiently validated. In the first article, we replicate results from the previous literature and test the methodology for robustness to event-induced variance and illiquidity. We conclude that existing methods for analysing the impact on bondholder wealth are intractable and that researchers should carefully consider sample characteristics, liquidity, and the presence of event-induced variance. The second part of this dissertation is devoted to the application of event study methodology in equity markets to test existing theoretical frameworks with regard to political and regulatory risk and uncertainty. The second and third articles focus specifically on the carbon risk hypothesis and examine whether carbon emissions are priced in by investors. We show that market reactions to regulatory announcements are driven significantly by greenhouse gas emissions. The results provide short-term evidence supporting the carbon risk hypothesis. In the fourth article, we analyse the market impact of populist success in national elections. We show that the increase in OHLC volatility in the run-up to an election is a robust proxy for the sensitivity of a stock to the election outcome. On the one hand, this dissertation thus advances research in the field of corporate bonds, and on the other hand, it contributes to the verification of existing theoretical models on the influence of regulatory and political risks and uncertainties on shareholder wealth.

Zusammenfassung

Natürliche Experimente haben in der empirischen Finanzforschung eine lange Tradition. Die Methodik der Ereignisstudien ist ein bewährter Ansatz zur Messung der wirtschaftlichen Auswirkungen eines Ereignisses auf das Vermögen der Aktionäre. Dahingehend umfasst diese Dissertation zwei Hauptthemenbereiche. Der erste Teil stellt einen methodischen Beitrag zur Grundlagenforschung im Bereich der Unternehmensanleihen dar. Die Entwicklung der Ereignisstudienmethodik steht hier weit hinter dem Stand der Ereignisstudienmethodik im Bereich der Aktien zurück. Bestehende Methoden sind zudem noch unzureichend validiert. Im ersten Artikel replizieren wir daher ein Simulationsexperiment aus der bestehenden Literatur und testen diese Ansätze auf ihre Robustheit gegenüber ereignisinduzierter Varianz und Illiquidität. Wir ziehen den Schluss, dass die bestehenden Methoden zur Analyse der Auswirkungen auf das Vermögen von Anleihegläubigern durchaus emfindlich gegenüber den beiden zuvorgenannten Phänomene sind und dass Forscher die Merkmale der Stichprobe, die Liquidität und das Auftreten von ereignisinduzierter Varianz sorgfältig berücksichtigen sollten. Der zweite Teil dieser Dissertation ist der Anwendung der Methodik der Ereignisstudien auf den Aktienmärkten gewidmet, um bestehende theoretische Rahmenwerke im Hinblick auf politische und regulatorische Risiken und Unsicherheiten zu testen. Der zweite und dritte Artikel konzentrieren sich speziell auf die Kohlenstoffrisikohypothese und untersuchen, ob Kohlenstoffemissionen von den Investoren eingepreist werden. Wir zeigen, dass die Marktreaktionen auf regulatorische Ankündigungen wesentlich von den Treibhausgasemissionen beeinflusst werden. Die Ergebnisse liefern Beweise für die Unterstützung der Kohlenstoffrisikohypothese. Im vierten Artikel analysieren wir die Auswirkungen des populistischen Wahlerfolgs auf die Finanzmärkte. Wir zeigen, dass der Anstieg der OHLC-Volatilität im Vorfeld einer Wahl ein robuster Indikator für die Empfindlichkeit einer Aktie gegenüber dem Wahlergebnis ist. Diese Dissertation bringt somit einerseits die Grundlagenforschung im Bereich der Unternehmensanleihen voran und trägt andererseits zur Überprüfung bestehender theoretischer Modelle über den Einfluss regulatorischer und politischer Risiken und Unsicherheiten auf das Vermögen der Aktionäre bei.

I thank all my great colleagues, comrades, and co-authors, many of whom became close friends during my academic journey. I also thank my supervisor, Dirk Schiereck, for his excellent guidance and strategic mentoring along the way.

to Mum, Dad, and Joe.

Contents

Chapter 1
Synopsis

"Like tipsy drinkers, financial markets are seeing the world upside down. Bad economic news is good news if it means more central bank action to support growth. Good news that means central banks may withdraw the punchbowl is bad news."
(Financial Times, 2013)

Financial markets have been mired in a state of uncertainty during the time of writing this thesis. The prolonged bull market that emerged in the aftermath of the 2008 financial crisis was abruptly halted with the onset of the COVID-19 pandemic in early 2020. This unprecedented event triggered a surge in uncertainty, resulting in a sharp decline in asset prices and increased volatility (see e.g., Baker et al., 2020; Pagano and Zechner, 2022) eventually prompting regulators and policymakers to intervene through regulatory measures (Baker et al., 2020; Boyarchenko et al., 2022). In addition, there are several geopolitical crises and tensions such as the trade war between the U.S. and China, the Russian invasion of Ukraine, the escalating tensions between China and Taiwan, the banking crisis in the United States, as well as the omnipresent threat of human-induced climate change (Cook et al., 2013; Haustein et al., 2017) severely affecting volatility in financial markets (see, e.g., Engle and Campos-Martins, 2023).

From a merely academic perspective, recent years offer a variety of significant events that provide ideal frameworks for natural experiments. Political decisions and related events continuously influence companies' risk and return profile and thus ultimately the valuation of the underlying assets (Schwert, 1981). The origins and channels are multifaceted and our understanding thereof is still limited. Asset prices are ultimately the result of buy and sell decisions that reflect the expected value of the future profitability of the underlying asset.[1] It is important to understand that actors in financial markets behave predictive and forward-looking, i.e., they anticipate probability-based expected values for the occurrence of certain events and their associated economic impact *ex-ante* (Schwert, 1981). Assuming that the efficient market hypothesis (EMH) holds and that "security prices at any point in time fully reflect all available information" (Fama, 1970, p. 383), researchers and practitioners can gain valuable insights into how the market evaluates the effectiveness of, for example, regulatory changes and policy decisions, *inter alia*. The market's estimate of these impacts should be exceptionally accurate, on average, albeit agnostic in the sense that one can only extract the aggregate net effect of all information penetrating the market, and it is empirically impossible to isolate the individual determinants.

The intricate relationship between politics and financial markets and the associated implications is a topic of considerable academic interest. This thesis contributes to the existing body of literature in this field. The first part provides a methodological contribution. While methods for event studies in equity markets are well established, comparable studies analysing the effects on bondholders' wealth are much less common. The second part is devoted to natural experiments on major political

[1] There is evidence that profitability is not the only determinant of share prices, e.g., if investors have a particular preference for certain types of assets and therefore derive non-pecuniary utility from holding them (see, e.g., Fama and French (2007) and Pástor et al. (2021))

events, analysing the resulting stock market reactions. In particular, we focus on regulatory carbon risk and geopolitical risk in the context of elections.

Methodological Challenges

The approach used in this paper is known as the "Event Study Methodology" (MacKinlay, 1997). The aim is to examine whether the occurrence of an event has significantly changed investors expectations about future profitability of a company. This should be reflected in a revaluation, which ultimately leads to abnormal stock returns, i.e., returns that deviate significantly from the estimates derived from an empirical asset pricing model, on average. Naturally, event study methodology is based on the assumption that financial markets are efficient. However, testing the EMH is empirically impossible because one inevitably faces the so-called joint hypothesis problem. Within the context of the EMH, the joint hypothesis problem refers to the difficulty of testing market efficiency independently from the respective asset pricing model employed (Fama, 1970, 1991). Every test for market efficiency is based on the assumption that the applied asset pricing model provides valid and reliable estimates and *vice versa*. It follows that if we observe anomalies in the behaviour of returns, it is ambiguous whether it is driven by a response of investors to new information or the result of an inadequate asset pricing model (Fama, 1970, 1991).

In contrast to equity-based research, however event study methodology in fixed-income research are far less established and insufficiently validated. The first part of this dissertation, therefore, deals with the methodological validation of existing approaches in fixed-income research. Specifically, we draw on the idea of Bessembinder et al. (2008) to incorporate bonds liquidity when examining the size and power of parametric and non-parametric tests. Furthermore, we follow Boehmer et al. (1991) and Marks and Musumeci (2017) in assessing the validity of exisiting approaches under the presence of event-induced variance. To this end, we replicate the empirical approach of Ederington et al. (2015) and incorporate the liquidity proxy λ, as proposed by Dick-Nielsen et al. (2012). Our results show that the power of existing approaches is sensitive to reduced sample sizes and increased noise. Moreover, researchers may account for liquidity, especially if a sample is tilted towards bonds with higher levels interest rate risk and credit risk.

Regulation and Legislation

Political decisions regarding regulations and legislation can have a profound effect on financial markets. Governments may implement rules and regulations to ensure stability, protect investors, and prevent market abuses. Changes in these regulations can impact market participants, their activities, and overall market dynamics (Schwert, 1981). Currently, one of the most prominent examples is the impact of (expected) more stringent regulations with regard to the sustainable transition, especially the regulation of carbon dioxide emissions.

Pástor et al. (2021) propose a two-factor model for investing that incorporates environmental, social, and governance (ESG) criteria. In this model, environmentally friendly (green) assets are characterized by low expected returns in equilibrium due to investors' preferences for holding them and their role as a hedge against climate risk. Pástor et al. (2021) assume that investors show an aversion to unexpected climate deterioration. Accordingly, assets associated with environmentally harmful practices (so-called brown assets) lose value compared to green assets in the event of an unforeseen climate deterioration, e.g., due to new government regulations that penalise brown companies. Thus far, there is some empirical evidence supporting the theoretical model of Pástor et al. (2021). Krueger et al. (2020) provide survey-based evidence that "institutional investors believe climate risks have financial implications for their portfolio firms and that these risks, particularly regulatory risks, already have begun to materialize." Stroebel and Wurgler (2021) confirm those findings, arguing that 861 finance academics, professionals, public sector regulators, and policy economists identify regulatory risk as the most salient climate risk for businesses and investors over the next five years. The empirical evidence on the carbon risk hypothesis in particular is, however, inconsistent. Bolton and Kacperczyk (2021), analyze the pricing of corporate-level CO2 emissions as an empirical risk factor in underlying asset prices. Nevertheless, scholars face difficulty in isolating carbon risk given the mechanical correlations between a firm's carbon emissions and firm fundamentals, especially in vendor-estimated data, and thus the exposure to incumbent risk factors (Aswani et al., 2024).

The second essay, Mueller et al. (2023b), considers the European Green Deal (EGD) in December 2019 to conduct a quasi-natural experiment. The announcement of the EGD differs from other green policy announcements. A five-page list of regulatory measures and restrictions was leaked twelve days before the official announcement so that endogeneity concerns, as well as the effect of investors' anticipations, are mitigated. The circumstance that the information was unexpectedly leaked at a specific point in time makes this event an ideal opportunity to distinguish carbon risk from traditional factor premiums. Bolton and Kacperczyk (2021) state that "if a large federal carbon tax were to be introduced, this would be a systematic shock affecting all companies with significant emissions." which should translate in significantly negative wealth effects conditional on firms' greenhouse gas emissions if carbon risk has not been priced in adequately *a priori*. Our results show that the overall market reaction of the European stock market was indeed negative and that greenhouse gas emissions and carbon emissions constitute an important determinant of abnormal returns. However, the results are moderate and induce that carbon risk has already been priced in, but somewhat underestimated, following the survey-based evidence of Stroebel and Wurgler (2021). We also find evidence of a noticeable correlation of the carbon premium with incumbent risk premia, supporting the findings of Aswani et al. (2024). Nevertheless, we insinuate that the findings in our natural experiment are attributable to carbon emissions rather than firm-fundamentals.

The third essay, Mueller et al. (2023c), reviews a similar research question. The 2022 reform of the European Emissions Trading System has been announced on Sunday, December 18, 2022. We document primarily negative market reactions, especially on the last trading day before the reform was announced. This can be viewed as a generic, undirected risk premium that investors require to bear the uncertainty associated with the expected announcement. Following the announcement, we observe primarily positive market reactions, which are positively related to the corporate-level emissions. Although the event considered in this study is indisputably less exogenous than the leakage of information regarding the EGD, our results pose important implications for future research. While prior research has documented mixed results on stock market reactions to green policy announcements (e.g., Ramiah et al. (2013)), we argue that it is imperative to consider investor anticipation in event studies, especially when it is known that there will be an event, but the outcome remains unclear *a priori*. The results of studies that analyze the average stock market response to multiple policy announcements simultaneously, e.g., Borghesi et al. (2022), should therefore be interpreted with utmost caution. Both aforementioned essays contribute to the ongoing discussion by providing short-term, event-induced evidence in support of the (regulatory) carbon risk hypothesis.

Geopolitical Events and Tensions

The U.S., Europe, and other Western democracies are witnessing a shift to the right and a strengthening of right-wing populist parties. Political events at the global level, such as elections, geopolitical tensions, trade disputes, or international conflicts, may induce uncertainty and, thus, volatility in financial markets (Pástor and Veronesi, 2012, 2013). For example, the announcement of new trade tariffs or the ascend of populism can substantially affect financial markets as investors assess the potential impact on global trade, supply chains, and economic growth (e.g., Wagner et al. (2018)).

Prediction and valuation of the associated economic impacts of election outcomes ex-ante is a challenge for researchers, especially in countries where coalition formation is uncertain after election results. Theoretical guidance on resulting stock market reactions is provided by Pástor and Veronesi (2012, 2013). Pástor and Veronesi (2012) propose a theoretical model of how the announcement of policy decisions and policy changes affect stock prices. Pástor and Veronesi (2013) go even further focusing on the way stock prices react to political signals about potential future policy decisions. The authors emphasize on the risk premium, volatility and correlation caused by political uncertainty. The models predict that equities are more volatile and more strongly correlated when political uncertainty is greater.

Wagner et al. (2018) offer empirical evidence, analyzing stock market reactions in response to the election of Donald J. Trump as the 45th President of the United States of America in November 2016. Stock market reactions were significantly related to the expected changes in tax and trade policies. Hanke et al. (2020) demonstrate the extent of market efficiency by exploiting changes in betting odds to infer expected changes in asset prices and *vice-versa*. These results illustrate how closely asset prices mirror policy developments and thereby serve as a gauge of the adequate estimation of expected and associated economic impacts.

The electoral success of far-right parties in Sweden and Italy in September 2022 provides the foundation for another quasi-natural experiment. The fourth article (Mueller et al., 2023a) exemplifies an investment strategy whereby investors are able to select stocks based on outcome-related sensitivity without the necessity of tedious data acquisition, selection, and modeling. We draw on the idea of Hanke et al. (2020) and leverage the hypothesis that policy changes and populist success influence volatility in financial markets (Hartwell, 2022; Pástor and Veronesi, 2013; Stöckl and Rode, 2021). The empirical approach of Hanke et al. (2020) is extended so that only historical intraday volatility (open-high-low-close, OHLC) derived from intraday stock prices is used to classify stocks, eliminating the need for any additional data. We show empirically that investors can hedge against the risks associated with elections (i.e., political uncertainty and impact risk in the notation of Pástor and Veronesi (2012)) by simply excluding stocks with the highest increase in intraday volatility in the run-up to the election.

Structure

The remainder of this thesis is organized as follows. Section 2 presents the first article (Mueller et al., 2024). Section 3 is devoted to the second article (Mueller et al., 2023b). Section 4 covers the third article (Mueller et al., 2023c). Section 5 exhibits the fourth article (Mueller et al., 2023a).

Chapter 2
Corporate Bond Market Event Studies: Event-Induced Variance and Liquidity

Abstract

This paper addresses the power of event studies in corporate bond markets. While an approach using standardized abnormal returns is well specified under standard conditions, we identify two market phenomena negatively impacting the informative value of results. In particular, we show that test power decreases rapidly in the presence of event-induced variance. Moreover, illiquidity becomes a material concern when the samples are geared towards above-average maturities and credit risks. Therefore, we suggest a refinement to the current standard approach and provide open-source tools to implement event studies.

This chapter is a working paper and has been published on SSRN. The version at the time of publication is attached in the appendix.

Mueller, L., Riehl, K., Buschulte, S., & Weiss, P. (2024). *Corporate bond market event studies: Event-induced variance and liquidity*, SSRN Working Paper No. 4859838. https://doi.org/10.2139/ssrn.4859838

Chapter 3
Is decarbonization priced in? – Evidence on the carbon risk hypothesis from the European Green Deal leakage shock

Abstract

On November 29, 2019, twelve days before the official announcement, information was leaked regarding the ambitions of the European Green Deal, i.e., the full decarbonization of the European Union by 2050 and lifting of 2030 emissions targets from 40% to 55%. The leakage should have triggered a Europe-wide systemic shock to financial markets without an accompanying announcement of supportive measures. Applying event study methodology to a sample of 600 European large and mid-cap stocks, we find that the overall market reaction was indeed significantly negative, albeit moderate. Abnormal returns gradually decline with increasing GHG emissions. Conversely, the official announcement emphasizing financial support and the green growth narrative did not ignite a positive market reaction. OLS regressions reveal that GHG emissions explain negative market reactions in response to the leak, whereas environmental performance and commitment are negatively related to returns obtained over intermediate horizons. We conclude that market participants incorporate available GHG emissions information into (short-term) reassessments following the promulgation of a significant environmental policy change.[1]

This chapter has been published as:

Mueller, L., Ringel, M., & Schiereck, D. (2023b). Is decarbonization priced in? - evidence on the carbon risk hypothesis from the European Green Deal leakage shock. *International Journal of Theoretical and Applied Finance*. https://doi.org/10.1142/S0219024923500188.

[1]We thank the participants and discussants of the Academy of Economics and Finance 2023, Liechtenstein Workshop on Sustainable Finance 2023, Conference on Climate and Energy Finance 2023, and Conference CSR, the Economy and Financial Markets 2023 for their valuable feedback.

Chapter 4
Regulatory Carbon Risk: Evidence from the 2022 Reform of the EU Emissions Trading Scheme

Abstract

This paper examines the market reaction of 600 European stocks to the announcement of the reform of the European Emissions Trading Scheme (ETS). We find significant negative CARs over the week before the announcement, yet firm-level GHG emissions, environmental performance, and other firm-specific controls fail to explain these. In contrast, we confirm a positive market response over the week following the announcement. Firm-level emissions and environmental performance are both positively associated with post-agreement CARs. What seems counterintuitive at first glance can be explained by the disparities between both metrics. From an investor's perspective, better environmental performance represents lower risk exposure to environment-related risk, regardless of the absolute level of externalities.

This chapter has been published as:

Mueller, L., Ringel, M., & Schiereck, D. (2023c). Regulatory Carbon Risk: Evidence from the 2022 reform of the EU Emissions Trading Scheme. *Zeitschrift für Umweltpolitk und Umweltrecht, 3/2023*[1]

[1]https://online.ruw.de/suche/zfu/Regulat-Car-Ris-Evide-fro-the-202-Ref-of-the-EU-Em-a19a5741d54785d34fbbdd337e149f6e

Chapter 5
Europe's gone „right" - A comparative study of stock market reactions to populist success in Sweden and Italy

Abstract

We analyze the reactions of national equity markets to the election of far-right populist governments in Italy and Sweden in September 2022. We apply event study methodology to samples of 285 Swedish and 144 Italian stocks. Share prices of Italian stocks largely aligned in the week before the vote. Conversely, the Swedish electoral outcome hit markets by surprise. Share prices adjusted in the days following the vote. We use firm-level increases in intraday volatility before the vote to estimate sensitivities to electoral outcomes. Dollar-neutral long-minus-short strategies based on these sensitivities prove explanatory power of historical intraday volatility in Sweden, where post-event reactions and uncertainty were evident. The model identified those stocks that are most sensitive to the election outcome.

This chapter has been published as:

Mueller, L., Bartel, M., & Schiereck, D. (2023a). Europe's gone 'right' – a comparative study of stock market reactions to populist success in Sweden and Italy. *Finance Research Letters, 55, Article: 103829.* https://doi.org/10.1016/j.frl.2023.103829

Bibliography

Aswani, J., Raghunandan, A., & Rajgopal, S. (2024). Are carbon emissions associated with stock returns? *Review of Finance*, *28*(1), 75–106. https://doi.org/10.1093/rof/rfad013

Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M., & Viratyosin, T. (2020). The unprecedented stock market reaction to COVID-19. *The Review of Asset Pricing Studies*, *10*(4), 742–758. https://doi.org/10.1093/rapstu/raaa008

Bessembinder, H., Kahle, K. M., Maxwell, W. F., & Xu, D. (2008). Measuring abnormal bond performance. *The Review of Financial Studies*, *22*(10), 4219–4258. https://doi.org/10.1093/rfs/hhn105

Boehmer, E., Musumeci, J., & Poulsen, A. B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, *30*(2), 253–272. https://doi.org/10.1016/0304-405X(91)90032-F

Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, *142*(2), 517–549. https://doi.org/10.1016/j.jfineco.2021.05.008

Borghesi, S., Castellini, M., Comincioli, N., Donadelli, M., Gufler, I., & Vergalli, S. (2022). European green policy announcements and sectoral stock returns. *Energy Policy*, *166*, 113004. https://doi.org/10.1016/j.enpol.2022.113004

Boyarchenko, N., Cox, C., Crump, R. K., Danzig, A., Kovner, A., Shachar, O., & Steiner, P. (2022). The primary and secondary corporate credit facilities. *Economic Policy Review*, *28*(1), 1–34. https://doi.org/10.2139/ssrn.4155392

Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs, P., & Skuce, A. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, *8*(2), 024024. https://doi.org/10.1088/1748-9326/8/2/024024

Dick-Nielsen, J., Feldhütter, P., & Lando, D. (2012). Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics*, *103*(3), 471–492. https://doi.org/10.1016/j.jfineco.2011.10.009

Ederington, L., Guan, W., & Yang, L. Z. (2015). Bond market event study methods. *Journal of Banking & Finance*, *58*, 281–293. https://doi.org/10.1016/j.jbankfin.2015.03.013

Engle, R. F., & Campos-Martins, S. (2023). What are the events that shake our world? measuring and hedging global COVOL. *Journal of Financial Economics*, *147*(1), 221–242. https://doi.org/10.1016/j.jfineco.2022.09.009

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

Fama, E. F. (1991). Efficient capital markets: Ii [Citations: 2,370]. *The Journal of Finance*, *46*(5), 1575–1617. https://doi.org/10.1111/j.1540-6261.1991.tb04636.x

Fama, E. F., & French, K. R. (2007). Disagreement, tastes, and asset prices. *Journal of Financial Economics*, *83*(3), 667–689. https://doi.org/10.1016/j.jfineco.2006.01.003

Hanke, M., Stöckl, S., & Weissensteiner, A. (2020). Political event portfolios. *Journal of Banking & Finance*, *118*, 105883. https://doi.org/10.1016/j.jbankfin.2020.105883

Hartwell, C. A. (2022). Populism and financial markets. *Finance Research Letters*, *46*, 102479. https://doi.org/10.1016/j.frl.2021.102479

Haustein, K., Allen, M. R., Forster, P. M., Otto, F. E. L., Mitchell, D. M., Matthews, H. D., & Frame, D. J. (2017). A real-time global warming index. *Scientific Reports*, *7*, 15417. https://doi.org/10.1038/s41598-017-14828-5

Krueger, P., Sautner, Z., & Starks, L. T. (2020). The importance of climate risks for institutional investors. *The Review of Financial Studies*, *33*(3), 1067–1111. https://doi.org/10.1093/rfs/hhz137

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, *35*(1), 13–39.

Marks, J. M., & Musumeci, J. (2017). Misspecification in event studies. *Journal of Corporate Finance*, *45*, 333–341. https://doi.org/10.1016/j.jcorpfin.2017.05.003

Mueller, L., Bartel, M., & Schiereck, D. (2023a). Europe's gone 'right' – a comparative study of stock market reactions to populist success in Sweden and Italy. *Finance Research Letters*, *55, Article: 103829*. https://doi.org/10.1016/j.frl.2023.103829

Mueller, L., Riehl, K., Buschulte, S., & Weiss, P. (2024). *Corporate bond market event studies: Event-induced variance and liquidity*, SSRN Working Paper No. 4859838. https://doi.org/10.2139/ssrn.4859838

Mueller, L., Ringel, M., & Schiereck, D. (2023b). Is decarbonization priced in? - evidence on the carbon risk hypothesis from the European Green Deal leakage shock. *International Journal of Theoretical and Applied Finance*. https://doi.org/10.1142/S0219024923500188

Mueller, L., Ringel, M., & Schiereck, D. (2023c). Regulatory Carbon Risk: Evidence from the 2022 reform of the EU Emissions Trading Scheme. *Zeitschrift für Umweltpolitk und Umweltrecht*, *3/2023*.

Pagano, M., & Zechner, J. (2022). COVID-19 and corporate finance. *The Review of Corporate Finance Studies*, *11*(4), 849–879. https://doi.org/10.1093/rcfs/cfac025

Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, *142*(2), 550–571. https://doi.org/10.1016/j.jfineco.2020.12.011

Pástor, L., & Veronesi, P. (2012). Uncertainty about government policy and stock prices. *The Journal of Finance*, *67*(4), 1219–1264. https://doi.org/10.1111/j.1540-6261.2012.01746.x

Pástor, L., & Veronesi, P. (2013). Political uncertainty and risk premia. *Journal of Financial Economics*, *110*(3), 520–545. https://doi.org/10.1016/j.jfineco.2013.08.007

Ramiah, V., Martin, B., & Moosa, I. (2013). How does the stock market react to the announcement of green policies? *Journal of Banking & Finance*, *37*(5), 1747–1758. https://doi.org/10.1016/j.jbankfin.2013.01.012

Schwert, G. W. (1981). Using financial data to measure effects of regulation. *The Journal of Law and Economics*, *24*(1). https://doi.org/10.1086/466977

Stroebel, J., & Wurgler, J. (2021). What do you think about climate finance? *Journal of Financial Economics*, *142*(2), 487–498. https://doi.org/10.1016/j.jfineco.2021.08.004

Stöckl, S., & Rode, M. (2021). The price of populism: Financial market outcomes of populist electoral success. *Journal of Economic Behavior & Organization*, *189*, 51–83. https://doi.org/10.1016/j.jebo.2021.06.037

Wagner, A. F., Zeckhauser, R. J., & Ziegler, A. (2018). Company stock price reactions to the 2016 election shock: Trump, taxes, and trade. *Journal of Financial Economics*, *130*(2), 428–451. https://doi.org/10.1016/j.jfineco.2018.06.013

Declaration of Honor

I declare upon my word of honor that the doctoral thesis submitted herewith is my own work. All sources and aids used have been listed. All references or quotations in any form and their usage have been clarified.

The dissertation has not been submitted for examination purposes to any institution before.

Appendix

Corporate Bond Market Event Studies: Event-Induced Variance and Liquidity*

Lukas Müller† Kevin Riehl‡ Sonja Buschulte§ Patrick Weiss¶

June 10, 2024

Abstract

This paper addresses the power of event studies in corporate bond markets. While an approach using standardized abnormal returns is well specified under standard conditions, we identify two market phenomena negatively impacting the informative value of results. In particular, we show that test power decreases rapidly in the presence of event-induced variance. Moreover, illiquidity becomes a material concern when the samples are geared towards above-average maturities and credit risks. Therefore, we suggest a refinement to the current standard approach and provide open-source tools to implement event studies.

Keywords:

Event study; Corporate bonds; Event-induced variance; Liquidity

JEL Codes:

C1, G10, G14

*We thank Dirk Schiereck as well as participants at Academy of Economics and Finance 2024, for their comments, insights, and discussion. We provide open-source code via this Github repository.

†Technical University of Darmstadt. E-mail: lukas.mueller2@tu-darmstadt.de.
‡Technical University of Darmstadt. E-mail: riehl.kevin@web.de.
§Technical University of Darmstadt. E-mail: buschulte.s@gmx.de.
¶Reykjavik University. E-mail: patrickw@ru.is.

1 Introduction

A common approach to measuring wealth effects in financial markets is to conduct an event study whereby researchers examine the residuals between realized returns and returns that are expected in the absence of the event. Naturally, any such approach is subject to the so-called joint hypothesis problem (Fama, 1970) as it is empirically impossible to test whether abnormal returns truly reflect market efficiency or are the result of flawed asset pricing models and biased estimates. Given that asset pricing models for equity markets are well-established and widely applied, existing research has focused primarily on the impact on shareholder wealth as a measure of the overall economic impact on firm value. The broad acceptance of equity pricing models is in stark contrast to corporate bond markets, where researchers do not commonly apply similar models. Nevertheless, bonds also represent claims on the same underlying cash flows yet with different payout profiles. Hence, if some events affect both expected profitability and the financial stability of a firm, bondholder wealth effects may be adverse to shareholder wealth effects. Therefore, studying bond returns provides more granular insights into investors' perceptions of event-induced changes in firm-level risk. If abnormal returns depend on the maturity of the underlying bond, this may entail additional information about the time horizon over which investors expect these risks to materialize economically.

In this paper, we take a comprehensive look at suggested approaches and issues in event studies using U.S. corporate bond data. While corporate bond markets are historically among the least transparent markets (Goldstein et al., 2007), data availability and quality increased dramatically with the introduction of the Trade Reporting and Compliance Engine (TRACE) system in the early 2010s. This introduction fuelled a growth in empirical studies on corporate bonds, but event studies are still comparatively scarce. Notably, existing studies are largely based on early TRACE data, i.e., data up to at most 2011 including the extreme outcomes of the financial crisis. However, the data availability and quality on corporate bonds has improved substantially thereafter. Therefore, questions arise about the validity of earlier results following these changes. Initially, we establish the baseline power of event studies for corporate bonds under general conditions. As a first contribution, we conduct artificial event studies with simulations using 10,000 iterations. Here, we draw from a new time period starting in 2013 to assess the validity and reliability of the existing approaches in most recent data. We find that the test power has increased substantially vis-à-vis earlier samples (e.g., the sample used in Ederington et al., 2015). Moreover, our results suggest that for non-parametric tests and a shock of 15 basis points (bp), a minimum sample size of about 100 firm×day observations is required to achieve sufficient test power, *ceteris paribus*. However, two common empirical phenomena in corporate bond markets negatively impact the power of this standard approach.

First, we find the power threshold to be sensitive to an increase in noise. On the one hand, we document that employing weekly returns, calculated as the price change between the first and the last available trading day of a week, results in lower test power despite an increase in sample size by 50%. On the other hand, extending the event window and, therefore, increasing the likelihood of contaminating returns with confounding information is not the only source of noise. It is well-documented in event studies from equity markets that certain events also increase residual variance, commonly referred to as event-induced variance (see, e.g., Boehmer et al., 1991). A key contribution of this article is to assess the power of the bond event study methodology under the presence of event-induced variance. For this purpose, instead of imposing a static shock of 15bp on abnormal returns to simulate events, we draw a random number from a normal distribution with a mean of 15bp and a standard deviation equal to the factorized standard deviation of the respective bond's return (similar to Marks & Musumeci, 2017, for equity markets). We progressively increase the disturbance factor from zero to three and find that test power decreases rapidly with increasing event-induced variance. While standardized abnormal returns appear to be the most robust, the test power drops to about 75% when the variance is doubled.

Second, bonds are traded much less frequently than equities, and illiquidity is a major concern of empirical bond research. Bessembinder et al. (2008) acknowledge that liquidity could be a potential source of misspecification but do not address this issue further. Although there is evidence that investors favor larger issues (see, e.g., Bao et al., 2011), the issue volume per se is a fairly crude indicator of corporate bond liquidity. Against this backdrop, we stratify test power conditional on bonds' liquidity, credit rating, and maturity. We employ various liquidity measures and confirm the effect for price-impact and transaction cost measures, specifically the measure λ, as proposed in Dick-Nielsen et al. (2012). Our results show that liquidity has a considerable impact on test power, especially when the samples are geared toward securities with higher credit risk and longer maturities. The effect, however, is most pronounced for the most illiquid bonds (i.e., the lowest liquidity quintile). We conclude that researchers may want to consider eliminating the most illiquid bonds from the sample to reduce noise and improve test power.

Methodologically, our paper is related to the extensive body of literature that examines the validity of the event study methodology based on simulation experiments. Brown and Warner (1985) conduct a simulation experiment in equity markets to assess the validity of asset pricing models and the probability of rejecting the true null hypothesis as well as the ability to detect abnormal performance. On the topic of variance, Boehmer et al. (1991) find that even with small increases in variance, the tests reject the true null hypothesis too often. The authors suggest standardizing abnormal returns to counter this issue. Marks and Musumeci (2017) show that the approach in Boehmer et al. (1991) is also robust to event-induced variance. Following the approaches in the equity-based literature, Bessembinder et al. (2008) conduct an initial simulation to assess the test power of studies of bond events based on daily and monthly bond returns. Subsequently, Ederington et al. (2015) revisit the validity and reliability of the bond event study methodology. *Inter alia*, the study finds that standardization of abnormal returns, as it has become common practice in equity-based event studies, significantly increases reliability. This is the starting point for our research. We evaluate the validity and reliability of the existing approaches for corporate bonds based on recent data and apply the methods of Boehmer et al. (1991) and Marks and Musumeci (2017) to assess their robustness to event-induced variance.

Additionally, we provide an open-source implementation of the presented methodology on Github, with detailed instructions on how to calculate various event-study-relevant abnormal return tables. This allows for easy replication of our work and, additionally, for a convenient utilization of corporate bond event study methodology.

2 Data and Methodology

2.1 Sampling

Our sampling procedure is aligned with the established literature on empirical corporate bond research using TRACE data (see, e.g., Dick-Nielsen et al., 2012). We proceed with the sampling as follows. First, we select all available bonds that are eligible for our analysis based on cross-sectional data provided in Mergent FISD and issued after 2000. We include bonds that are senior, unsecured, non-asset-backed, and non-defeased while excluding securities with secured lease obligations or security pledges. We limit the sample to the most common bond types: U.S. Corporate Debentures, U.S. Corporate MTN (Medium Term Note), U.S. Corporate MTN Zero, U.S. Corporate Zero, and U.S. Corporate Bank Note. We exclude bonds issued by foreign ("yankee") and Canadian issuers. The remaining bonds are exclusively USD-denominated with fixed or zero coupon types. Bonds under Rule 144A, private placements, defaulted bonds, and those with given filing dates or settlement dates are excluded. We further remove putable bonds along with convertible, pay-in-kind, exchangeable, preferred, and perpetual bonds from the sample. This procedure yields 67,537 bonds, of which 46,355 bonds have trades reported in TRACE. We source and clean this data as outlined in Dick-Nielsen (2009).[1]

In addition, we apply several filters to the TRACE data to mitigate the effects of outliers and reduce potential noise. Specifically, we include only trades with a volume of minimum $50,000 to exclude retail-sized trades[2] and remove any rows containing missing values in yield to maturity, reported price, or entered volume quantity. Subsequently, we focus on trades within the price range of $25 to $1,000. We further remove observations where the reported

[1] We use the code provided by Scheuch et al. (2023) to filter Mergant FISD data and clean TRACE data.

[2] It has become common practice to exclude trades with a trading volume of less than $100,000 when working with TRACE data. However, this eliminates a considerable fraction of trades and, consequently, trading days. Although this certainly benefits computational efficiency, we propose a less restrictive threshold and include trades with a minimum trading volume of $50,000. Results for alternative thresholds are presented later in the paper.

price deviates by more than 10% from the daily median. Finally, we keep all observations with maturities within the range of one year to a maximum of 50 years. We use the rating time series from Standard and Poor's, Moody's, and Fitch. If no rating from Moody's is available, we use ratings from Standard and Poor's if available and Fitch otherwise. We remove all observations where no rating of any of the three agencies is available. In the end, the final sample covers 6,777,096 unique bond×day observations from 13,483 individual bonds and 1,275 individual firms in our sample period from 2013 until 2022.

We also compute quarterly liquidity measures that are regularly used in corporate bond research, following Dick-Nielsen et al. (2012). In particular, we use the Amihud measure, Roll measure, firm and bond zero trading days, internal round trip cost (IRC), turnover, and Amihud and IRC risk. Their construction and descriptive statistics are detailed in Appendix A. We use λ as our main liquidity proxy, which is calculated as the standardized sum of the Amihud measure (Amihud, 2002), the IRC measure (Feldhütter, 2012) and their respective standard deviations. Furthermore, we also consider a principal component decomposition of these individual liquidity measures to verify λ's appropriateness (see Appendix A).

2.2 Abnormal Returns

We calculate the daily bond price $\mathrm{P}_{n,t}$ for bond n on day t as the trade-weighted average of the trade prices P_{n,t_i} of all trades $t_i \in T_{n,t}$ on that day with the trade volume quantity N_{n,t_i} denoting the weights, i.e.,

$$\mathrm{P}_{n,t} = \frac{\sum\limits_{t_i \in T_{n,t}} \mathrm{P}_{n,t_i} \times \mathrm{N}_{n,t_i}}{\sum\limits_{t_i \in T_{n,t}} \mathrm{N}_{n,t_i}}. \tag{1}$$

We proceed with calculating unadjusted two-days returns $\mathrm{R}_{n,t}$ centered at time t as

$$\mathrm{R}_{n,t} = \frac{\mathrm{P}_{n,t+1} - \mathrm{P}_{n,t-1}}{\mathrm{P}_{n,t-1}}. \tag{2}$$

For the sake of simplicity, we use the reported prices as Bessembinder et al. (2008) note that accrued interest has a negligible influence on the results. Table 1 reports descriptive statistics of the bond- and firm-level returns. Panel A reports descriptive statistics on raw returns, while Panel B reports winsorized returns. The additional sample statistics in Table 2 show that the sample's properties and the characteristics of our bonds are very similar to those of Ederington et al. (2015).

- TABLES 1 AND 2 HERE -

We calculate three different abnormal returns on the bond level n, following Ederington et al. (2015): (i) abnormal returns $\mathrm{ABR}_{n,t}$, (ii) standardized abnormal returns $\mathrm{SABR}_{n,t}$, and (iii) abnormal standardized returns $\mathrm{ABSR}_{n,t}$. Abnormal returns are calculated as the residual between return $\mathrm{R}_{n,t}$ and benchmark return $\mathrm{BMR}_{n,t}$, as outlined in Equation (3). The benchmark return for bond n on day t is calculated as an equally-weighted average over all bond returns on day t that share the same rating and maturity group with bond n on this day. For forming the benchmark groups, we consider four maturity groups (i.e., zero to three, more than three to five, more than five to ten, and more than ten years) and six rating groups (i.e., Aaa and Aa, A, Baa, Ba, B, and below B). We determine benchmark returns only if the sample of bonds sharing the same rating and maturity on this day is greater than or equal to five. We compute firm-level abnormal returns $\mathrm{ABR}_{f,t}$ as the average over the respective bond-level abnormal returns available on a given day.

$$\mathrm{ABR}_{n,t} = \mathrm{R}_{n,t} - \mathrm{BMR}_{n,t} \tag{3}$$

The standardized abnormal returns are calculated by dividing the abnormal returns by their standard deviation, i.e.,

$$\mathrm{SABR}_{n,t} = \frac{\mathrm{ABR}_{n,t}}{\sigma_{n,t,ABR}}, \tag{4}$$

where $\sigma_{n,t,ABR}$ represents the standard deviation of (unshocked) abnormal returns over the period of $[t-55, t-6]$ and $[t+6, t+55]$.

7

The abnormal standardized returns $\text{ABSR}_{n,t}$ are calculated as the difference of standardized raw returns $\text{SRR}_{n,t}$ and standardized benchmark $\text{SBM}_{n,t}$ as outlined in Equation (5). $\text{SRR}_{n,t}$ denotes standardized raw returns, meaning $\text{R}_{n,t}$ divided by $\sigma_{n,t,R}$, where $\sigma_{n,t,R}$ is the standard deviation of returns $\text{R}_{n,t}$ over the period of $[t-55, t-6]$ and $[t+6, t+55]$. The standardized benchmark $\text{SBM}_{n,t}$ is the equally-weighted average of all $\text{SRR}_{n,t}$ for all bonds that share the same rating and maturity group (applying the same group definitions as before) with bond n on that specific day t. Again, we determine the standardized benchmark returns $\text{SBM}_{n,t}$ only if the sample of bonds sharing the same rating and maturity on this day is greater than or equal to five.

$$\begin{aligned} \text{SRR}_{n,t} &= \frac{\text{R}_{n,t}}{\sigma_{n,t,R}} \\ \text{ABSR}_{n,t} &= \text{SRR}_{n,t} - \text{SBM}_{n,t} \end{aligned} \tag{5}$$

Note, besides $\text{ABSR}_{n,t}$ that is based on standard deviations over the period from $[t-55, t-6]$ and $[t+6, t+55]$, we also calculate $\text{ABSR\_pre}_{n,t}$ which considers standard deviations calculated only over the pre-event period $[t-101, t-6]$. This ensures that we only use available information and avoid a look-ahead bias. Finally, we consider $\text{ABSR}_{n,t}$, $\text{SABR}_{n,t}$, and $\text{ABSR\_pre}_{n,t}$ which are winsorized at the 1% level.

2.3 Size and Power Tests

We examine three statistical tests for the presence of an event effect: the standard parametric t-test (abbreviated t, Student, 1908), the non-parametric Wilcoxon signed-rank test (abbreviated SR, Wilcoxon, 1947), and the non-parametric sign test (abbreviated S, Conover, 1999). Results are reported at the 1% significance level (two-sided), and we also report some results additionally at the 5% significance level.

To assess the accuracy of these three tests, we conduct simulations with 10,000 iterations.

In particular, we select 300 firm×date observations at random per trial. Where possible, we calculate the abnormal returns for the firms' traded bonds as outlined above. Based on this set, we evaluate the two accuracy measures.

Size test: Under the null hypothesis, each of the 300 abnormal returns (residuals) is normally distributed with a mean $\mu = 0$ and variance σ^2. If a statistical test erroneously rejects the null hypothesis, it is classified as a type 1 error (false positive). Over our 10,000 iterations, the fraction of cases in which the tests incorrectly rejected the null hypothesis is referred to as the "size test".

Power test: In the second test, we shock the 300 randomly drawn observations by inducing an event effect. In particular, we induce artificial shocks $\xi_{n,t,R}$ to the returns $\mathrm{R}_{n,t}$, with $\xi_{n,t,R}$ representing a normally distributed variable $\xi_{n,t,R} \sim \mathcal{N}(\mu, \sigma_n^2)$. These shocks come in two specifications. First, in the case of static shocks $\sigma_n = 0$ and $|\mu| = 15$bp. Second, when testing for event-induced variance, we follow Marks and Musumeci (2017) and induce stochastic shocks, i.e., $\sigma_n = \sigma_{n,t,R}$ and $|\mu| = 15$bp, where $\sigma_{n,t,R}$ is the standard deviation of raw bond returns $\mathrm{R}_{n,t}$ over the period of $[t-55, t-6]$ and $[t+6, t+55]$. Cases in which a test does not correctly detect an event (i.e., the test does not reject the null hypothesis), are referred to as type 2 errors (false negative). Overall, the fraction of cases in which the tests correctly rejected the null hypothesis is referred to as the "power test".

3 Baseline Results

In this section, we establish baseline results for our core contributions. On the one hand, we show results for unstandardized abnormal returns. On the other hand, we provide insights into standardized returns based on the general heteroskedasticity of bond returns. In both parts, we use the simulation procedure described before based on a random set of 300 firm×date observations, which we repeat 10,000 times and average across all experiments.

3.1 Abnormal Bond Returns

The point of departure for bond event studies is based on abnormal returns. As defined in Equation (3), abnormal returns adjust returns by subtracting the respective benchmark's returns. The benchmark matches the bond's two main risk factors: Credit and interest rate risk. We use this simulation to establish a baseline against which other specifications can be assessed to gain insights into the tests' size and power. Therefore, we also include financial firms and use a trade-volume cut-off of $50,000, below which a trade is excluded. Table 3 presents the first benchmark. Results obtained employing alternative thresholds are provided as robustness checks in Section 4.

- TABLE 3 HERE -

We start with assessing the type 1 error rate (i.e., the size test) of the three statistical tests. A test is well specified based on the significance level of 1% (5%) if the non-event null-rejection rate is below 0.5% (2.5%) in each tail. Based on a simulation of 10,000 trials, the *true* null hypothesis must not be rejected more than 100 (500) times. The results presented in Panel A of Table 3 suggest that the three tests (i.e., t-test, signed-rank test, and sign test) are well specified in terms of size. In most cases, the non-event null-rejection rates are below the respective significance level. The only exception is the signed-rank test, which tends to provide false evidence of an event. In particular, the type 1 error rate reaches the critical level of 0.5% in the upper tail (positive events, 1% significance level) and 2.52% at the lower tail (negative events, 5% significance level). On the other hand, our findings provide a more favorable result for the signed-rank test, as the size test values are much closer to the critical threshold than results of Ederington et al. (2015), who documented a type 1 error rate of 3.12% for negative events.

Table 3's Panel B shows the results of the power tests. As described before, we induce artificial shocks of 15bp to simulate an event and check whether the tests correctly reject the null hypothesis. The test power for positive events (+15bp) and negative events (-15bp) is evaluated separately. The power of the *t*-test to correctly detect positive events at 54.43% (at the 5% significance level) is insufficient, although it is considerably higher compared to the results of Ederington et al. (2015) (20.62%). Put differently, the null hypothesis of the *t*-test is only rejected in 5,443 out of 10,000 simulations based on 300 events. The power for identifying negative events is comparably low at 52.97%. Naturally, the results based on the 1% significance level are considerably lower at approximately 30%. The non-parametric tests, on the other hand, perform substantially better in our sample. Even without further adjustments, these tests exhibit sufficient power in detecting events. Moreover, we document negligible differences in power between tests for positive and negative events. The power of both the signed-rank and the sign test rises well above 95% at the 5% significance level, whereas Ederington et al. (2015) report much lower power ranging from approximately 50% to 70%.

In line with Marks and Musumeci (2017), we examine whether winsorization substantially impacts our results. As shown in Panel B of Table 1, controlling for outliers considerably reduces the excess kurtosis and skewness. We document an increase in the test power of the *t*-test to approximately 40% (1% significance level) while the test size still remains within the required threshold. Naturally, neither the size nor the power of the non-parametric tests are affected by winsorization.

3.2 Standardized Abnormal Bond Returns

This section addresses the issue of heteroskedasticity of bond returns. Bonds are primarily subject to credit and interest rate risk. Prior literature (see, e.g., Bessembinder et al., 2008) has shown that the variance of bond returns increases with higher levels of credit and interest rate. Therefore, we present the standard deviation of the ABR conditional on credit and interest rate risk in Table 4.

- TABLE 4 HERE -

Higher credit and interest rate risk result in larger standard deviations, which confirms the presence of heteroskedasticity. The only exception is the class of non-investment grade bonds, although this is likely to be driven by the limited sample size (as reported in Panel B below). To encounter the issue of heteroskedasticity, Ederington et al. (2015) propose standardizing abnormal returns as it is common practice in equity-based event study methodology. Consequently, we establish a second baseline for standardized returns to contrast the already reasonable test power we find, especially when applying the non-parametric tests.

- TABLE 5 HERE -

Table 5 Panel A shows the return characteristics of the two main standardized abnormal return measures. The level of kurtosis obtained with ABSR and SABR demonstrates that standardization reduces the sample's heteroskedasticity beyond the effect of winsorization. The size tests performed in Panel B show that for standardized ABR, all three tests are also reasonably well specified. However, we also find that the tests exceed the thresholds in some cases, especially considering positive events and the signed-rank test. This is consistent with earlier results without standardization. The power tests in Panel C show that standardization of ABR substantially increases test power. This increase in power is robust across the three different ways of standardization, i.e., ABSR, SABR, and ABSR-Pre. However, based on these findings, we recommend using ABSR or SABR over ABSR-Pre. Standardization provides the greatest improvement in the power of the t-test, but non-parametric tests continue to be more powerful at power levels of 94% and above. Finally, Panel D illustrates that the improvement in explanatory power through standardization is more pronounced for moderate shocks of 10bp than for larger shocks of 25bp. We conclude that standardization of ABR is imperative in bond event studies, especially regarding events with potentially moderate effects.

In Table 6, we depict the test power of bivariate-sorted ABSR conditional on the level of credit and interest rate risk. The categories with higher credit quality bonds (basically all investment grade bonds) and maturity up to ten years exhibit superior test power of nearly 100% for all groups. While the test power remains close to 100% for AAA to A rated bonds with maturities over ten years, it drops to 91.36% for BAA rated bonds and to 69.26% and 68.52% for BA and B rated bonds, respectively. Test power for non-investment grade bonds is relatively low across all maturity buckets. However, the sample sizes of these groups, as denoted in Panel B, are limited and, thus, inevitably also contribute to the reduction in test power.

- TABLE 6 HERE -

4 Sampling and Sample Size

In this section, we investigate whether the more restrictive sample selection drives the documented increase in test power. In addition, we examine the effects of sample size on test size and power. In the last section, we provide results based on weekly returns. We limit ourselves to verbatim explanations to avoid excessive repetition of similar results and show tabulated results in Appendix B.

4.1 Alternative Sampling

There are two essential sampling decisions where researchers take different routes in bond event studies. On the one hand, the decision to exclude retail-sized trades differs. On the other hand, researchers sometimes exclude financial firms from the sample. We investigate the impact of those choices on test size and power.

In some cases, event studies do not exclude potentially noisy retail-sized trades from their main sample (prominently, e.g., Ederington et al., 2015). In contrast, we exclude trades with less than \$50,000 in volume from our main sample, and other studies exclude trades below \$100,000. Any filter like this results in a lower number of available firm×day observations, potentially reducing test power due to the decrease in sample size. However, lower levels of noise should theoretically increase test power, *ceteris paribus*. We find that incorporating retail-sized trades decreases the size of the signed-rank test to 0.57. As expected, the test power (at the 1% significance level) of the t-test decreases marginally by $1 - \frac{31.78\%}{34.12\%} = 6.86\%$ ($1 - \frac{29.27\%}{32.16\%} = 9.00\%$ for negative events). However, counter-intuitively, the power of non-parametric tests even increases. These results suggest that excluding retail-sized trades does not significantly improve the power of the test. Secondly, the effect of an increased sample size appears to outweigh the benefits of (potential) noise reduction. Nevertheless, the results show that excluding retail-sized trades does not drive the higher test power *vis-à-vis* Ederington et al. (2015). These implications are further substantiated when turning to the results for the sample based on trades with a volume above \$100,000. As expected, the signed-rank test's size improves slightly, eventually dropping below the statistical threshold of 0.5%. There is, however, no considerable gain in power regarding the t-test, and, in line with the trend previously identified, the power of the non-parametric tests decreases slightly.

It is also common practice to exclude financial companies from empirical analyses. Financial companies carry exposure to systemic risk and are subject to a complex regulatory environment (Fama & French, 1992). In addition, financial companies and the financial market often affect each other, and results could be distorted by effects of financial market change, e.g., interest rate factors, instead of clearly showing the response to a specific event (Foerster & Sapp, 2005). Hence, we assess the impact of restricting the sample to non-financial firms. We exclude all stocks with a first-digit SIC code of 6, i.e., finance, real-estate, and insurance firms. Excluding bonds issued by financial institutions reduces the sample size, eventually causing a marginally lower test power.[3]

Overall, the decision as to whether financial companies or retail-sized are included or excluded has no material impact on the results. If researchers run the risk of reducing the sample size excessively (e.g., due to the potential exclusion of illiquid assets, as shown later in section 6), we advocate the inclusion of retail-sized trades based on our findings as opposed to the established convention of removing trades below $100,000.

4.2 Sample Size

The results of the previous section indicate that test power is relatively sensitive to changes in sample size. We conduct another series of simulations to answer the question of what the minimum sample size should be and how sensitive the results are to a reduction in sample size. Specifically, we run the simulations as specified before, but adjust the number of random draws from 100 to 1,500 bond$\times$day for each set of 10,000 rounds. Naturally, we see a steady increase in the effective sample size from approximately 25 to 400 actual returns. In particular, with 300 random draws, the average number of calculable abnormal returns is approximately 80, resulting in a coverage ratio of 26.67%. The ratio remains comparable for 25 out of 100 and 400 out of 1,500 draws. The result of this analysis is a graph of test size and test power as a function of sample size for each of the three considered tests, which we present in Figure 1.

- FIGURE 1 HERE -

The results of the size tests, shown in Panel A of Figure 1, are largely robust in showing no dependence on the sample size and, thereby, corroborate the tabulated results. Again, the signed-rank test shows a tendency towards type 1 errors, especially for negative events. On the other hand, the test power strongly depends on the sample size, as illustrated in Panel B of Figure 1. While the test power of the two non-parametric tests increases exponentially with sample size and eventually saturates in a range between 100 and 150 observations,

[3] The average number of abnormal returns computable from a random draw of 300 bond$\times$day observations decreases slightly from 80.32 to 79.40.

the test power of the t-test increases much more slowly. In fact, the t-test's power remains well below a sufficient level, even at a sample size of 400. The graphs also illustrate that an average sample size of 80, on which the results of this paper are based, approaches the minimum sample size required for bond event studies based on the estimated impact of 15bp. Overall, our results suggest that for non-parametric tests, a minimum sample size of about 100 firm×day observations is required to achieve sufficient test power, *ceteris paribus*.

4.3 Alternative Event Window Length

The length of the event window varies from study to study, and there is no uniform, generalized approach. While rather uncommon in equity studies, some fixed-income studies propose to use weekly returns to increase the sample of included firms. However, this comes at the cost of incorporating potentially confounding information, thereby increasing the noise level. Hence, we tackle this issue empirically and compute our results for returns calculated on a weekly basis. The calculation of returns is analogous to Equation (2) except that we calculate weekly returns between the first and last day of a week on which a bond was traded. An additional requirement is the existence of at least one trading day in between, such that we calculate at least the two-day returns as shown in Equation (2). We deliberately choose the maximum return period possible to demonstrate the impact of larger levels of natural noise on the test size and power. Arguably, in an actual event study, choosing the minimum return period could be advisable to reduce the likelihood of including potentially confounding information. However, as some information may be leaked or anticipated *a priori*, it is often necessary to expand the return period. To illustrate the effects of noise and to obtain conservative results, we use the former approach. We show the results of the size and power tests on weekly returns in Table 7.

- TABLE 7 HERE -

While most return characteristics remain comparable, the total number of observations decreases when transitioning to weekly returns. However, despite that, the average sample size per draw increases by approximately 50% to 120. As our earlier results show, this should theoretically lead to an increase in test power. However, Panel C of Table 7 indicates that the three tests in our analysis exhibit considerably lower test power based on weekly returns. Furthermore, as shown in Panel D of Table 7, the reduction in test power is more pronounced for events with a moderate magnitude. In conclusion, it should be carefully balanced whether potential advantages, e.g., the inclusion of a larger number of companies in the sample and extending the event period, outweigh the potential downside, i.e., incorporating confounding information resulting in a higher level of noise. This is particularly concerning in the scenario of event-induced variance, which we explore in Section 5.

5 Event-induced Variance

Overall, our previous results provide evidence that the existing methodology of event studies in corporate bond markets is reliable under standard conditions. However, our findings also indicate that these results are generally susceptible to increases in noise and variance. Hence, in this section, we address an issue prevalent in equity-based event studies but which has yet to be addressed in bond-based event studies. In fact, several events are known to cause an increase in the variance of stock returns (see, e.g., Brown et al., 1988) and Marks and Musumeci (2017) tested equity-based event studies for event-induced variance. Following the equity-based literature, we thus extend the existing methodology for bond-event studies by testing whether the methodology is also robust to event-induced variance.

We repeat the simulations as specified above to investigate event-induced variance in bond-event studies with the necessary adjustments. Instead of adding a static shock of ±15bp, we follow Marks and Musumeci (2017) by adding a random number ξ drawn from a normal distribution with a mean of ±15bp and a standard deviation equal to the standard deviation of the returns used to calculate the ABSR, i.e., $\xi_{n,t} \sim \mathcal{N}(\mu,\, \sigma^2_{n,t})$. For the size test, we add a number drawn from a similar distribution in terms of standard deviation, albeit

17

with a mean of zero, i.e., $\xi_{n,t} \sim \mathcal{N}(0,\, \sigma^2_{n,t})$. We choose the induced variance (i.e., $\sigma^2_{n,t}$) based on the available data instead of forcing an arbitrary number. Otherwise, we use the same procedure using 300 bond×day draws for 10,000 trials with two-day returns.

- TABLE 8 HERE -

Panel B of Table 8 shows that all three tests are comparably well specified in terms of size. However, the power of the applied tests decreases with the introduction of event-induced variance. For example, the test power of the t-test is diminished from approximately 90% to 57% for ABSR and 68% for SABR. This reduction in test power comes despite the added standardization. Albeit still significantly reduced, the test power of the signed-rank test based on SABR stands out among the tests at around 75% for both positive and negative shocks. Turning to alternative magnitudes, as shown in Panel D of Table 8, we find that the ABSR-based test remains well specified for shocks with a magnitude of 25bp on average.

Following these findings, we illustrate a series of power curves to assess the sensitivity of the applied tests to event-induced variance. Figure 2 plots the test power conditional on the level of event-induced variance. Specifically, we run the experiments as above, but the level of event-induced variance is increased incrementally by a factor γ. The induced shock ξ is thus a number drawn from a normal distribution $\xi_{n,t} \sim \mathcal{N}(\mu,\, \sigma^2)$, where $\sigma^2 = \gamma \times \sigma^2_{n,t}$.

- FIGURE 2 HERE -

Figure 2 shows that the test power decreases with an increase in event-induced variance. A dramatic reduction in test power can be recognized consistently over all panels, even with small increases in variance. The vertical lines indicate γ=0.5 and γ=1. Interestingly, SABR appear superior to the other approaches, especially under the signed-rank test, in line with the results presented in Table 8. We conclude from this analysis of event-induced variance that although the test power of ABSR and SABR differs only slightly under standard conditions, the experiment suggests the use of SABR when an event is expected to cause a substantial increase in variance.

6 The Role of Liquidity

It is a stylized fact that liquidity in corporate bond markets is substantially lower than in equity markets. Consequently, there is a wealth of empirical research on the impact of liquidity on bond prices and returns (see, e.g., Goldstein & Namin, 2023, for an overview). Nevertheless, the effect of liquidity on test size and power in event studies has yet to be thoroughly addressed.[4] However, given its importance, we investigate liquidity's role explicitly.

In the analysis of liquidity's effects, we focus on λ (Dick-Nielsen et al., 2012) as the main liquidity measure .[5] λ is calculated as the standardized sum of the Amihud measure (Amihud, 2002), the internal round trip cost (IRC) measure (Feldhütter, 2012)) and their respective standard deviations. In addition, we provide details on the construction in Appendix A.

- TABLES 9 AND 10 HERE -

To understand the liquidity's impact on returns, we first stratify the standard deviation and sample size of abnormal bond returns by our primary liquidity proxy λ and both credit rating (see Table 10) and time to maturity (see Table 9), respectively. We find clear patterns between liquidity and the two other risk dimensions. Specifically, bonds with a remaining time to maturity of one to three years are much more liquid than bonds with a maturity of more than ten years, as indicated by the return variability and the number of observations per portfolio. This pattern also applies to portfolios sorted by credit rating and liquidity. Independent of the other dimensions, the standard deviation itself increases monotonically with decreasing liquidity, which is robust across all maturity ranges and rating categories. Overall, we clearly see that liquidity has a sizeable impact on the heteroskedasticity of bond returns, comparable to credit and interest rate risk.

- TABLES 11 AND 12 HERE -

Next, given the impact on heteroskedasticity, we investigate the impact of liquidity on

[4] In fact, Bessembinder et al. (2008) report that they formed alternative benchmark portfolios using issuance size as a proxy for liquidity risk but conclude that it does not improve test statistics. However, issuance size is only a crude liquidity proxy.

[5] Alternatively, we use the Roll measure and arrive at the same conclusions.

inference from event studies using the same simulation protocol as before. However, we specifically focus on the most powerful test from our previous analyses, i.e., we use ABSRs with the signed-rank test. Indeed, Tables 11 and 12 show that liquidity substantially impacts test power. In general, the test power decreases with lower liquidity, as well as with longer maturities and higher credit risk. In particular, the impact of liquidity is most pronounced for bonds with higher levels of interest rate and credit risk. For the most extreme portfolios, the test power reaches levels that are considered insufficient, e.g., 56.06% for the most illiquid bonds with a remaining maturity of more than ten years. Regarding the credit risk dimension, the results are also remarkable. For example, the test power for BA-rated bonds falls to 73% in the lowest liquidity quantile.[6]

One salient finding in this exercise is the implicit correlation between the liquidity of corporate bonds and credit risk, which is already well-established in the existing literature (see, e.g., Helwege et al., 2014). After stratifying by credit risk, sorting by liquidity does not substantially change the number of observations in each portfolio. This is in stark contrast to the bi-variate sorting by maturity and liquidity. However, although the sample sizes in the former cases remain relatively stable (or even increase), the test power is considerably lower for portfolios with low liquidity. This points to a non-linear moderating effect between liquidity spreads and credit spreads in line with the theoretical predictions in Ericsson and Renault (2006). Finally, these results imply that corporate bond liquidity should be considered in corporate bond event studies, especially when samples are geared towards above-average maturities and higher levels of credit risk.

[6] We arrive at virtually identical results if we use the Roll measure as a proxy for liquidity, which appears intuitive given a correlation of 0.619 between both measures (see Table A-2 in Appendix A). For example, the test power for the least liquid bonds in the five to ten year maturity range falls to 83.47% and for bonds with maturities over ten years to 64.09%. When stratified by credit rating and liquidity, the results are similar. The test power for the least liquid falls to 81.70% and 69.73% for BAA- and BA-rated bonds, respectively. We do not observe such distinct patterns when using turnover as a proxy for liquidity. The results are tabulated in Appendix B.

7 Conclusion

In this paper, we examine the validity and reliability of event study methodology in corporate bond markets. We find that the empirical approach using standardized abnormal returns jointly with non-parametric tests (popularized by Ederington et al., 2015) is well-specified. Standardizing ABR and using non-parametric tests are superior to the use of raw ABR and parametric tests in terms of test size and power. However, a sufficient sample size is crucial to obtain reliable results. Based on artificial shocks of 15bp, the non-parametric tests converge to a reasonable power level at a sample size of about 100, all else equal.

In addition, we assess the validity of the methodology to a common problem in event studies, i.e., event-induced variance. We document that test power diminishes rapidly with increasing variance. Moreover, similar conclusions arise when evaluating the test power based on weekly returns. We find that constructing weekly returns (i.e., returns based on the first and last available trading day of the week) leads to an increase in noise, which reduces the test power substantially.

Finally, illiquidity is a prominent issue in corporate bond markets. Thus, we evaluate the sensitivity of test power to bond liquidity. We show that liquidity effects are particularly pronounced for lower-rated bonds and bonds with longer maturities. From this, we conclude that researchers need to assess the characteristics of their sample carefully. Statistical inference is challenging due to severe power issues in samples with longer maturity bonds or higher credit risk. Therefore, excluding the most illiquid assets can reveal if higher noise levels drive the lack of a finding.

Figures

Figure 1: Test size and power conditional on sample size. This figure displays the results of 16 simulations (10,000 repetitions each) based on ABR with increasing sample size from 100 random firm×day draws until 1,500 draws. For the power tests, we induce static, positive shocks (i.e., $\xi_{n,t,R}$) with $|\mu| = 15$bp and $\sigma = 0$. With 100 (1,500) firm×day observations drawn at random, the average sample yields 25 (400) available returns. Panel A reports the results of the size tests, and Panel B reports the results of the power tests. We present the results of the standard parametric t-test (t), the non-parametric Wilcoxon signed-rank test (SR), and the non-parametric sign test (S).

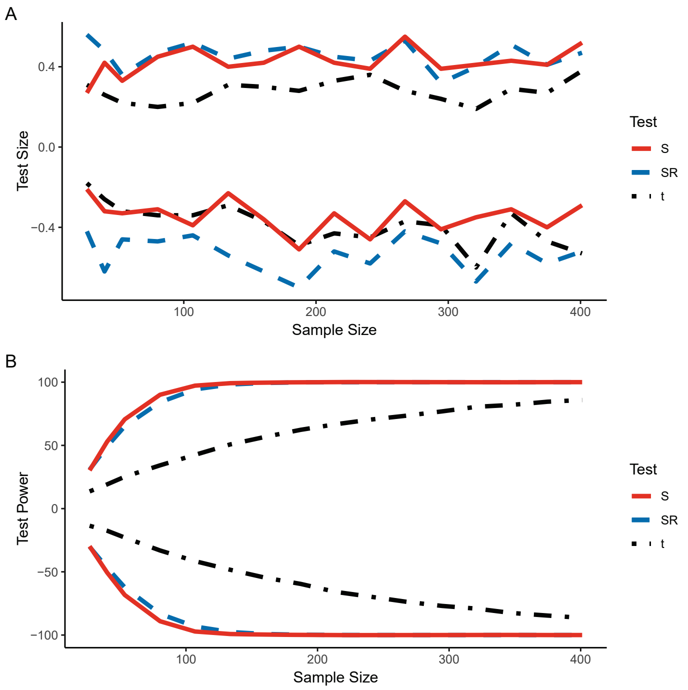
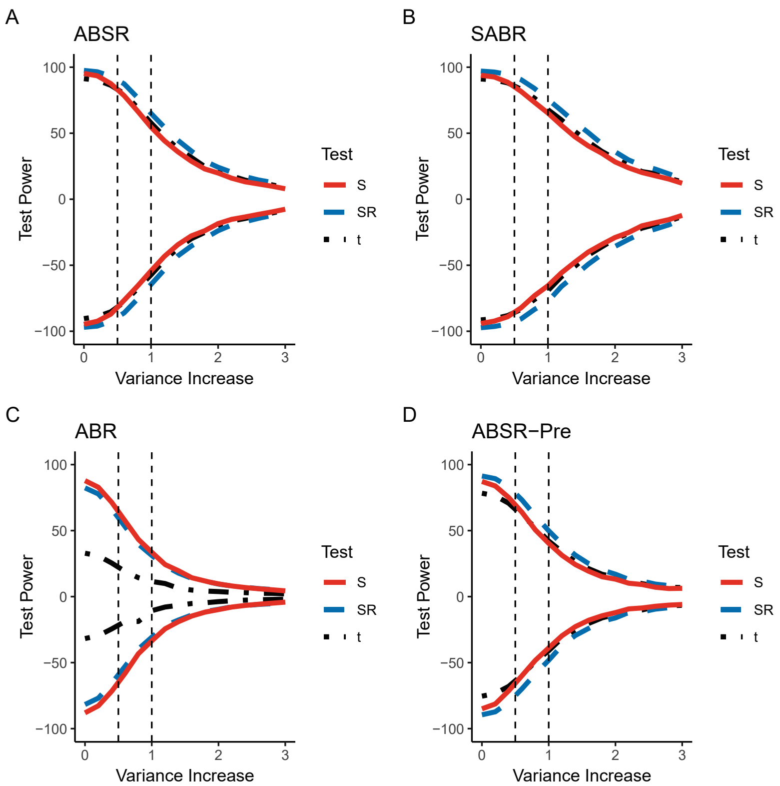


Figure 2: Test power conditional on event-induced variance. This figure shows the results of 60 simulations (10,000 repetitions of 300 firm×day random draws each). We incrementally increase the standard deviation (i.e, σ) of the positive shock (i.e., $\xi_{n,t,R}$) with $|\mu| = 15$bp. Formally, σ is calculated as $\gamma \times \sigma_{n,t,R}$, where $\sigma_{n,t,R}$ is the standard deviation of raw bond yields (i.e., $R_{n,t}$ during the periods $[t-55, t-6]$ and $[t+6, t+55]$) and γ ranges from zero to three. Panel A shows results for ABSR, Panel B for SABR, Panel C for ABR, and Panel D for ABSR-Pre. We present the results of the standard parametric t-test (t), the non-parametric Wilcoxon signed-rank test (SR), and the non-parametric sign test (S).



23

Tables

This page was intentionally left blank.

Table 1: Descriptive statistics of the bond- and firm-level returns. This table presents descriptive statistics on the bond- and firm-level returns. Panel A states descriptives of raw bond returns, abnormal returns (ABR), and firm-level ABR based on TRACE data from 2013 to 2022. Raw (two-day) bond returns are based on trade-weighted daily prices (see Equation (1)), and ABR are defined as the two-day bond return minus the average return of all bonds in the same rating/maturity group. The firm-level ABR are calculated as the trade volume-weighted average of the firm's ABR. Panel B reports the statistics after winsorization at the 1 and 99% levels.

Panel A: Returns

	Mean	Median	SD	Skew	Kurt	% positive	N	% days
Raw bond returns	-0.0014	0.0008	1.0659	-0.7352	160.37	49.7743	3,054,101	9.2261
Abnormal bond returns	0.0000	-0.0018	0.7778	-0.0368	267.88	49.6975	3,038,955	9.1796
Abnormal firm returns	0.0004	-0.0002	0.8220	-1.4739	294.02	49.9636	857,451	25.8166

Panel B: Winsorized Returns

	Mean	Median	SD	Skew	Kurt	% positive
Raw bond returns	0.0003	0.0007	0.7732	-0.0674	3.4467	49.7743
Abnormal bond returns	0.0005	-0.0018	0.5478	0.0411	4.0409	49.6975
Abnormal firm returns	0.0023	-0.0002	0.5473	0.011	4.7398	49.9636

Table 2: Descriptive statistics of the bond sample. This table presents descriptive statistics on the corporate bonds used throughout this paper. Panel A shows additional bond characteristics (i.e., the maturity, the rating, the liquidity measure λ, the offering amount (in millions), and the coupon). Panel B shows the distribution of the sample bonds based on credit rating and maturity.

Panel A: Sample Statistics

	Mean	Median	SD
Maturity (in years)	9.61	6.18	8.87
Rating	14.38	15.00	3.14
λ (Liquidity)	0.14	-0.77	2.99
Offering Amount ($\times 10^6$)	709	500	673
Coupon	3.45	2.60	3.78

Panel B: Share per Rating and Maturity

Rating	Share	Maturity	Share
AAA to AA	7.42%	1-3 years	19.93%
A	30.38%	3-5 years	21.48%
BAA	41.81%	5-10 years	37.45%
BA	12.41%	> 10 years	21.14%
B	6.59%		
Below B	1.39%		

Table 3: Size and power tests using unstandardized abnormal returns. This table presents the size and power tests using firm-level abnormal returns (ABR) at a significance level (SL) of 1% and 5% (two-sided), respectively. The tests include the parametric t-test (t), the non-parametric signed-rank test (SR), and the sign test (S). For each test, we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws. Panel A presents the size tests. In Panel B, we show the power tests for artificial positive and negative 15bp shocks.

Panel A: Size Tests

	t		SR		S	
SL	0.5	99.5	0.5	99.5	0.5	99.5
Size	0.19	0.25	0.47	0.50	0.43	0.33
SL	2.5	97.5	2.5	97.5	2.5	97.5
Size	1.86	2.22	2.52	2.39	1.89	1.91

Panel B: Power Tests - 15bp Shocks

	SL	t	SR	S
Positive shock	0.5	34.96	85.5	90.26
Negative shock	99.5	32.16	83.68	89.31
Positive shock	2.5	54.43	95.24	97.33
Negative shock	97.5	52.97	94.59	97.02

Table 4: Abnormal returns' standard deviations stratified by rating and maturity. This table presents the standard deviations of bond-level abnormal returns (ABR) over six rating and four maturity cut-offs. Panel A states the standard deviation in percent. Panel B reports the corresponding number of observations.

Panel A: Standard Deviation

	1-3 years	3-5 years	5-10 years	> 10 years
AAA to AA	0.2094	0.2894	0.4149	0.6998
A	0.2186	0.3104	0.4379	0.7807
BAA	0.3937	0.4957	0.6589	1.0350
BA	0.7011	0.8791	0.9188	1.3149
B	1.7895	1.4452	1.3421	1.9983
Below B	2.5697	2.4344	2.2795	1.9799

Panel B: Number of Observations

	1-3 years	3-5 years	5-10 years	> 10 years
AAA to AA	53,944	47,866	68,481	55,894
A	242,491	206,695	268,560	209,840
BAA	230,547	246,898	466,972	332,057
BA	52,701	87,841	200,493	37,242
B	19,965	51,496	119,653	5,037
Below B	3,930	12,691	17,559	102

Table 5: Size and power tests of standardized abnormal returns. This table presents size and power tests of firm-level abnormal standardized returns (ABSR), standardized abnormal returns (SABR), unstandardized abnormal returns (ABR, raw), and ABSR with pre-event period $[t-101, t-6]$ standard deviation (ABSR-Pre). The tests include the parametric t-test (t), the non-parametric signed-rank test (SR), and the sign test (S). For each test, we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws. Panel A shows descriptive statistics of returns in percent. Panel B presents the test size using a two-sided significance level (SL) of 1%. Panel C shows the power tests for artificial positive and negative 15bp shocks. Panel D repeats the power tests for alternative shocks of 10bp and 25bp.

Panel A: Return Characteristics

	Mean	Median	SD	Skew	Kurt	N
ABSR	0.0003	-0.0036	0.8093	0.0478	2.5283	2,860,243
SABR	-0.0021	-0.0064	1.0682	0.0432	1.9244	2,862,094

Panel B: Size Tests

	t		SR		S	
SL (%)	0.5	99.5	0.5	99.5	0.5	99.5
ABSR	0.40	0.51	0.41	0.54	0.37	0.30
SABR	0.41	0.51	0.55	0.62	0.56	0.36
ABR (raw)	0.19	0.25	0.47	0.50	0.43	0.33
ABSR-Pre	0.42	0.65	0.39	0.69	0.26	0.43

Panel C: Power Tests - 15bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
ABSR	89.97	96.97	94.77	91.97	97.70	95.38
SABR	91.07	96.93	93.99	92.23	97.35	93.94
ABR (raw)	32.16	83.68	89.31	34.96	85.50	90.26
ABSR-Pre	75.74	89.69	85.67	78.51	91.30	87.13

Panel D: Power Tests - 10bp and 25bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
10bp - raw	12.37	46.44	56.18	14.18	49.96	56.41
10bp - stand.	55.00	75.56	71.24	58.87	77.45	71.50
25bp - raw	71.27	99.53	99.81	73.31	99.65	99.88
25bp - stand.	99.91	99.98	99.91	99.96	99.98	99.91

Table 6: Test power stratified by bond rating and maturity for a 15bp shock. This table reports the test power of the signed-rank test at the 1% significance level based on bond-level abnormal standardized bond returns. We group bonds by six rating and four maturity cut-offs. The test power in Panel A is the average test power across positive and negative 15bp shocks (i.e., we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws). Panel B reports the corresponding number of observations.

Panel A: Power Test - 15bp Shocks

	1-3 years	3-5 years	5-10 years	> 10 years
AAA to AA	100.00	100.00	100.00	99.29
A	100.00	100.00	100.00	98.04
BAA	100.00	100.00	100.00	91.36
BA	100.00	100.00	99.39	69.26
B	99.57	99.70	97.72	68.52
Below B	64.28	78.89	77.65	22.91

Panel B: Number of Observations

	1-3 years	3-5 years	5-10 years	> 10 years
AAA to AA	50,582	46,140	65,972	53,335
A	225,423	197,073	255,026	196,448
BAA	210,776	232,166	439,092	309,991
BA	49,503	84,092	192,095	35,476
B	18,382	48,745	113,723	4,721
Below B	3,340	11,558	16,496	88

Table 7: Size and power tests based on weekly returns. This table presents size and power tests of firm-level abnormal standardized returns (ABSR), standardized abnormal returns (SABR), unstandardized abnormal returns (ABR, raw), and ABSR with pre-event period $[t-101, t-6]$ standard deviation (ABSR-Pre). The tests include the parametric t-test (t), the non-parametric signed-rank test (SR), and the sign test (S). For each test, we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws. Panel A shows descriptive statistics of returns in percent. Panel B presents the test size using a two-sided significance level (SL) of 1%. Panel C shows the power tests for artificial positive and negative 15bp shocks. Panel D repeats the power tests for alternative shocks of 10bp and 25bp.

Panel A: Return Characteristics

	Mean	Median	SD	Skew	Kurt	N
ABSR	0.0008	-0.0074	0.8628	0.0765	2.5927	510,428
SABR	-0.0111	-0.0161	1.2424	0.0429	1.6431	511,680

Panel B: Size Tests

	t		SR		S	
SL (%)	0.5	99.5	0.5	99.5	0.5	99.5
ABSR	0.25	0.34	0.49	0.36	0.35	0.34
SABR	0.52	0.21	0.66	0.26	0.45	0.17
ABR (raw)	0.24	0.29	0.45	0.44	0.42	0.38
ABSR-Pre	0.32	0.54	0.41	0.55	0.42	0.39

Panel C: Power Tests - 15bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
ABSR	75.42	87.53	82.17	77.31	87.97	80.49
SABR	81.91	89.38	80.49	78.22	86.37	75.85
ABR (raw)	20.51	71.93	82.21	23.04	75.69	82.62
ABSR-Pre	73.35	87.79	82.26	76.75	90.18	84.16

Panel D: Power Tests - 10bp and 25bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
10bp - raw	7.17	34.13	44.19	9.1	38.24	45.63
10bp - stand.	38.64	55.42	49.94	40.8	55.25	46.81
25bp - raw	55.65	98.63	99.61	57.18	99.07	99.55
25bp - stand.	98.98	99.67	98.89	99.15	99.68	98.7

Table 8: Size and power tests with event-induced noise. This table presents size and power tests of firm-level abnormal standardized returns (ABSR), standardized abnormal returns (SABR), unstandardized abnormal returns (ABR, raw), and ABSR with pre-event period $[t-101, t-6]$ standard deviation (ABSR-Pre). The tests include the parametric t-test (t), the non-parametric signed-rank test (SR), and the sign test (S). For each test, we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws, where we introduce event-induced variance equal to the standard deviation used to scale ABSR. Panel A shows descriptive statistics of returns in percent. Panel B presents the test size using a two-sided significance level (SL) of 1%. Panel C shows the power tests for artificial positive and negative 15bp shocks. Panel D repeats the power tests for alternative shocks of 10bp and 25bp.

Panel A: Return Characteristics

	Mean	Median	SD	Skew	Kurt	N
ABSR	0.0003	-0.0036	0.8093	0.4780	2.5283	2,860,243
SABR	-0.0021	-0.0064	1.0682	0.0432	1.9244	2,862,094

Panel B: Size Tests

	t		SR		S	
SL (%)	0.5	99.5	0.5	99.5	0.5	99.5
ABSR	0.50	0.47	0.51	0.49	0.32	0.39
SABR	0.54	0.50	0.44	0.59	0.38	0.46
ABR (raw)	0.30	0.34	0.56	0.50	0.41	0.37
ABSR-Pre	0.36	0.43	0.33	0.48	0.24	0.36

Panel C: Power Tests - 15bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
ABSR	56.77	64.12	54.11	57.23	64.78	54.81
SABR	67.53	74.94	64.91	67.97	75.61	64.87
ABR (raw)	11.11	29.84	32.69	10.78	30.97	33.39
ABSR-Pre	39.26	46.18	37.10	42.41	50.00	40.70

Panel D: Power Tests - 10bp and 25bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
10bp - raw	4.08	11.73	12.63	4.24	12.22	12.63
10bp - stand.	22.80	28.23	22.74	24.77	30.32	23.75
25bp - raw	33.98	73.85	77.73	35.81	75.61	79.76
25bp - stand.	95.80	97.11	92.46	95.96	97.15	92.37

Table 9: Abnormal returns' standard deviations stratified by liquidity and maturity. This table presents the standard deviations of bond-level abnormal returns (ABR) over five liquidity and four maturity cut-offs. Panel A states the standard deviation in percent. Panel B reports the corresponding number of observations.

Panel A: Standard Deviation

	1-3 years	3-5 years	5-10 years	> 10 years
High	0.2032	0.3107	0.4003	0.6914
2	0.2906	0.3914	0.4952	0.6458
3	0.3858	0.5150	0.5922	0.7153
4	0.6195	0.7060	0.7924	0.9001
Low	1.7031	1.3337	1.2205	1.5439

Panel B: Number of Observations

	1-3 years	3-5 years	5-10 years	> 10 years
High	240,883	147,370	190,021	52,603
2	148,469	146,278	219,339	116,953
3	115,543	130,976	229,788	154,567
4	84,838	147,245	240,439	158,218
Low	33,095	114,294	312,038	171,428

Table 10: Abnormal returns' standard deviations stratified by liquidity and rating. This table presents the standard deviations of bond-level abnormal returns (ABR) over five liquidity and six rating cut-offs. Panel A states the standard deviation in percent. Panel B reports the corresponding number of observations.

Panel A: Standard Deviation

	AAA to AA	A	BAA	BA	B	Below B
High	0.2305	0.2762	0.3582	0.4857	0.5995	0.8059
2	0.3222	0.3724	0.4562	0.5449	0.7248	1.1035
3	0.4137	0.4495	0.5703	0.6853	0.8757	1.5025
4	0.5301	0.5531	0.7597	0.9323	1.2279	1.7292
Low	0.6644	0.7788	1.2184	1.3047	2.1980	3.2007

Panel B: Number of Observations

	AAA to AA	A	BAA	BA	B	Below B
High	43,443	224,474	278,847	56,073	23,700	4,340
2	46,048	207,127	247,757	86,274	38,403	5,430
3	51,391	197,900	249,010	89,655	38,352	4,566
4	48,445	176,798	260,970	96,556	40,909	7,062
Low	34,713	107,378	293,044	117,854	62,998	14,868

Table 11: Test power stratified by liquidity and maturity for a 15bp shock. This table reports the test power of the signed-rank test at the 1% significance level based on bond-level abnormal standardized bond returns. We group bonds by five liquidity and four maturity cut-offs. The test power in Panel A is based on positive 15bp shocks (i.e., we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws). Panel B reports the corresponding number of observations.

Panel A: Power Test - 15bp Shocks

	1-3 years	3-5 years	5-10 years	> 10 years
High	100.00	100.00	100.00	73.69
2	100.00	100.00	99.96	92.40
3	100.00	100.00	99.88	90.63
4	100.00	100.00	99.43	83.54
Low	100.00	98.78	93.76	56.06

Panel B: Number of Observations

	1-3 years	3-5 years	5-10 years	> 10 years
High	235,539	153,166	199,605	54,641
2	152,539	150,718	226,129	116,520
3	122,888	136,508	233,988	150,889
4	95,494	149,134	244,082	152,793
Low	48,169	117,939	308,043	164,496

Table 12: Test power stratified by liquidity and rating for a 15bp shock. This table reports the test power of the signed-rank test at the 1% significance level based on bond-level abnormal standardized bond returns. We group bonds by five liquidity and six rating cut-offs. The test power in Panel A is based on positive 15bp shocks (i.e., we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws). Panel B reports the corresponding number of observations.

Panel A: Power Test - 15bp Shocks

	AAA to AA	A	BAA	BA	B	Below B
High	100.00	100.00	100.00	99.84	98.18	99.78
2	100.00	100.00	100.00	99.69	97.64	81.18
3	100.00	100.00	100.00	96.45	87.11	12.79
4	100.00	100.00	99.95	89.18	63.30	22.65
Low	99.97	100.00	93.25	73.00	42.12	12.70

Panel B: Number of Observations

	AAA to AA	A	BAA	BA	B	Below B
High	44,307	224,463	281,098	61,881	26,133	5,069
2	46,923	210,276	254,065	88,795	40,038	5,809
3	51,893	200,621	254,356	91,550	40,754	5,099
4	49,396	180,731	264,759	97,954	41,297	7,366
Low	35,616	115,133	293,836	118,016	62,138	13,908

References

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56. https://doi.org/10.1016/S1386-4181(01)00024-6

Bao, J., Pan, J., & Wang, J. (2011). The illiquidity of corporate bonds. *Journal of Finance*, *66*(3), 911–946. https://doi.org/10.1111/j.1540-6261.2011.01655.x

Bessembinder, H., Kahle, K. M., Maxwell, W. F., & Xu, D. (2008). Measuring abnormal bond performance. *The Review of Financial Studies*, *22*(10), 4219–4258. https://doi.org/10.1093/rfs/hhn105

Boehmer, E., Masumeci, J., & Poulsen, A. B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, *30*(2), 253–272. https://doi.org/10.1016/0304-405X(91)90032-F

Brown, K. C., Harlow, W. V., & Tinic, S. M. (1988). Risk aversion, uncertain information, and market efficiency. *Journal of Financial Economics*, *22*(2), 355–385. https://doi.org/10.1016/0304-405X(88)90075-X

Brown, S. J., & Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics*, *14*(1), 3–31. https://doi.org/10.1016/0304-405X(85)90042-X

Conover, W. J. (1999). *Practical nonparametric statistics* (Vol. 350). John Wiley & Sons.

Dick-Nielsen, J. (2009). Liquidity biases in trace. *The Journal of Fixed Income*, *19*(2), 43–55. https://doi.org/10.3905/jfi.2009.19.2.043

Dick-Nielsen, J., Feldhütter, P., & Lando, D. (2012). Corporate bond liquidity before and after the onset of the subprime crisis. *Journal of Financial Economics*, *103*(3), 471–492. https://doi.org/10.1016/j.jfineco.2011.10.009

Ederington, L., Guan, W., & Yang, L. (2015). Bond market event study methods. *Journal of Banking & Finance*, *58*, 281–293. https://doi.org/10.1016/j.jbankfin.2015.03.013

Ericsson, J., & Renault, O. (2006). Liquidity and credit risk. *The Journal of Finance*, *61*(5), 2219–2250. https://doi.org/10.1111/j.1540-6261.2006.01056.x

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Feldhütter, P. (2012). The same bond at different prices: Identifying search frictions and selling pressures. *The Review of Financial Studies*, *25*(4), 1155–1206. https://doi.org/10.1093/rfs/hhr093

Foerster, S. R., & Sapp, S. G. (2005). Valuation of financial versus non-financial firms: A global perspective. *Journal of International Financial Markets, Institutions and Money*, *15*(1), 1–20. https://doi.org/10.1016/j.intfin.2004.01.003

Friewald, N., Jankowitsch, R., & Subrahmanyam, M. (2012). Illiquidity or credit deterioration: A study of liquidity in the us corporate bond market during financial crises. *Journal of Financial Economics*, *105*(1), 18–36. https://doi.org/10.1016/j.jfineco.2012.02.001

Goldstein, M. A., Hotchkiss, E. S., & Sirri, E. R. (2007). Transparency and liquidity: A controlled experiment on corporate bonds. *The Review of Financial Studies*, *20*(2), 235–273. https://doi.org/10.1093/rfs/hhl020

Goldstein, M. A., & Namin, E. S. (2023). Corporate bond liquidity and yield spreads: A review. *Research in International Business and Finance*, *65*, 101925. https://doi.org/10.1016/j.ribaf.2023.101925

Helwege, J., Huang, J.-Z., & Wang, Y. (2014). Liquidity effects in corporate bond spreads. *Journal of Banking & Finance*, *45*, 105–116. https://doi.org/10.1016/j.jbankfin.2013.08.018

Marks, J. M., & Musumeci, J. (2017). Misspecification in event studies. *Journal of Corporate Finance*, *45*, 333–341. https://doi.org/10.1016/j.jcorpfin.2017.05.003

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, *39*(4), 1127–1139. https://doi.org/10.1111/j.1540-6261.1984.tb03897.x

Scheuch, C., Voigt, S., & Weiss, P. (2023). *Tidy finance with r* (1st). Chapman; Hall/CRC. https://doi.org/10.1201/b23237

Student. (1908). The probable error of a mean. *Biometrika*, *6*(1), 1–25. https://doi.org/10.2307/2331554

Wilcoxon, F. (1947). Probability tables for individual comparisons by ranking methods. *Biometrics*, *3*(3), 119–122. https://doi.org/10.2307/3001946

Appendix

A Liquidity Measures

In this section, we outline the calculation of the individual corporate bond liquidity measures drawing on the procedure of Dick-Nielsen et al. (2012). We also provide some descriptive statistics on the derived liquidity measures as presented in the original article. Again, we follow the structure of the original paper to maintain comparability. We also reproduce the descriptive statistics, the correlation matrix and the principal component analysis as presented in the original paper.

The first measure of (ill-)liquidity we consider is the Amihud measure (Amihud, 2002), which is defined as the price impact of a trade per unit traded. Formally, the Amihud measure is derived by calculating the daily average of the absolute returns r_j divided by the trading volume Q_j of the consecutive transactions of a given bond per day:

$$Amihud_t = \frac{1}{N_t}\sum_{j=1}^{N_t}\frac{|r_j|}{Q_j} = \frac{1}{N_t}\sum_{j=1}^{N_t}\frac{|\frac{P_j - P_{j-1}}{P_{j-1}}|}{Q_j} \tag{A-1}$$

Conceptually, it is necessary for a bond to trade at least twice per day to calculate the Amihud measure. We derive quarterly measures by taking the median of all daily observations in a quarter.

The bid-ask spread refers to the difference between the highest price a buyer is willing to pay (bid price) and the lowest price a seller is willing to accept (ask price) for a particular financial instrument. For bid-ask spreads, we employ two distinct proxies: the Roll measure (Roll, 1984) and Imputed Roundtrip Trades (IRC) as defined by Feldhütter (2012). Roll (1984) finds that, given certain assumptions, the percentage bid-ask spread is equal to two times the square root of minus the covariance between successive returns:

$$Roll_\tau = 2\sqrt{-cov(R_i, R_{i-1})} \tag{A-2}$$

Where τ is the period for which the measure is calculated. If the covariance is negative, the observation is discarded. The underlying rationale is that the bond price bounces between the bid and ask prices, and that higher percentage bid-ask spreads lead to higher negative covariance between successive yields (Dick-Nielsen et al., 2012). We define a monthly roll measure if a bond trades at least ten times in this month. We then define a quarterly rolling measure by taking the mean of the daily measures within the quarter.

Feldhütter (2012) presents an alternative method for assessing transaction costs, which he terms "Imputed Roundtrip Trades" (IRT). In corporate bond markets, it is not uncommon to observe a bond being traded multiple times in quick succession after a prolonged period of inactivity. This typically occurs when a dealer facilitates a transaction between a buyer and a seller and collects the bid-ask spread as compensation. When a match is found, two trades take place - one between the seller and the dealer and another between the buyer and the dealer. In some cases, if the matching involves a second dealer, there may also be a transaction between the two dealers. When two or three trades of the same bond with identical trade sizes occur on the same day, and there are no other trades with the same size, we consider these transactions as part of an Imputed Roundtrip Trade. For an IRT, the Imputed Roundtrip Cost (IRC) is defined as follows:

$$IRC = \frac{P_{\max} - P_{\min}}{P_{\max}} \tag{A-3}$$

where $P_{\max}$ represents the highest price within the IRT, and $P_{\min}$ is the lowest price within the IRT. To estimate daily roundtrip costs, the average of roundtrip costs for different trade sizes on that day is calculated. Quarterly IRC measures are the average of all daily estimates per bond and quarter.

Friewald et al. (2012) differentiate between trading activity measures and liquidity measures. Three volume-based activity measures that are considered are quarterly turnover of bonds as well as bond and firm zero trading days. The former is calculated as the cumulative trading volume of a bond per quarter divided by its issue size, both measured in \$. The latter is calculated as the number of actual trading days per bond or firm and quarter relative to total the number of trading days per quarter.

Dick-Nielsen et al. (2012) argue that investors are likely to consider not only the absolute level of bond liquidity, but also the possible future level and consequently the variability of the liquidity. Similar reasoning applies to returns of corporate bonds, as risk premia may depend on both level and risk of a bond's liquidity. Hence, we follow Dick-Nielsen et al. (2012) and also include the quarterly average standard deviations of the daily Amihud measure and the assumed round-trip costs.

In the initial paper, the authors conduct linear correlation analysis and principal component analysis of all liquidity proxies to see if most of the relevant information in the liquidity proxies can be captured by a few factors. Based on their results, they then construct a factor that loads equally on all four variables, i.e., Amihud, IRC, and their standard deviations, which they call λ. In Table A-1, we exhibit the results of the same analysis based on our sample. The results are virtually identical to the results presented in the original study. The cumulative explained variance of the first three components is 39.8%, 58.1%, and 71.2%, respectively. Dick-Nielsen et al. (2012) report very similar values of 39%, 59%, and 72%, respectively. The loadings of the first component also paint a very similar picture as in Dick-Nielsen et al. (2012). The first component is dominated by Amihud, IRC, and Roll, whereas the second component is driven by volume-based measures such as Bond Zero and Turnover.

Based on these findings we proceed with calculating our primary liquidity variable λ, formally denoted as:

$$\lambda_{it} = \sum_{j=1}^{4} \widetilde{L}_{it}^{j} \tag{A-4}$$

We provide descriptive statistics of all liquidity proxies in Table A-2. The results are virtually identical to Dick-Nielsen et al. (2012). Percentiles of the liquidity variables (Panel A) as well as their correlation (Panel B) are comparable in magnitude for the recent period of

2013 to 2022. Furthermore, we plot the liquidity variables' time series in Figure 3.

Table A-1: PCA of liquidity measures. This table provides the results of a Principal Component Analysis (PCA) of the derived liquidity measures similar to Dick-Nielsen et al. (2012). The first principal component correlates with the Amihud, IRC, and Roll (price impact) measures, whereas the second component represents a volume-based measure, i.e., turnover.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Amihud	0.431	-0.133	0.146	0.052	-0.338	0.800	-0.022	0.132
Roll	0.409	-0.297	-0.002	-0.495	0.060	-0.250	-0.647	0.121
Firm Zero	-0.025	-0.292	-0.840	0.171	-0.416	-0.068	-0.025	-0.009
Bond Zero	-0.143	-0.657	-0.115	-0.121	0.593	0.229	0.301	0.161
TO	0.089	0.560	-0.470	-0.530	0.280	0.292	0.113	-0.030
IRC	0.460	-0.005	-0.115	0.407	0.395	0.019	-0.112	-0.664
Amihud Risk	0.448	-0.144	0.106	-0.333	-0.261	-0.324	0.669	-0.190
IRC Risk	0.454	0.209	-0.116	0.389	0.234	-0.224	0.129	0.682
CEV (%)	39.8	58.1	71.2	79.8	87.8	93.3	97.3	100.0

Table A-2: Statistics of Liquidity Proxies. This table presents statistics of liquidity proxies for corporate bonds. The proxies are calculated quarterly for each bond. Panel A displays percentiles for the different proxies, while Panel B presents the correlations matrix.

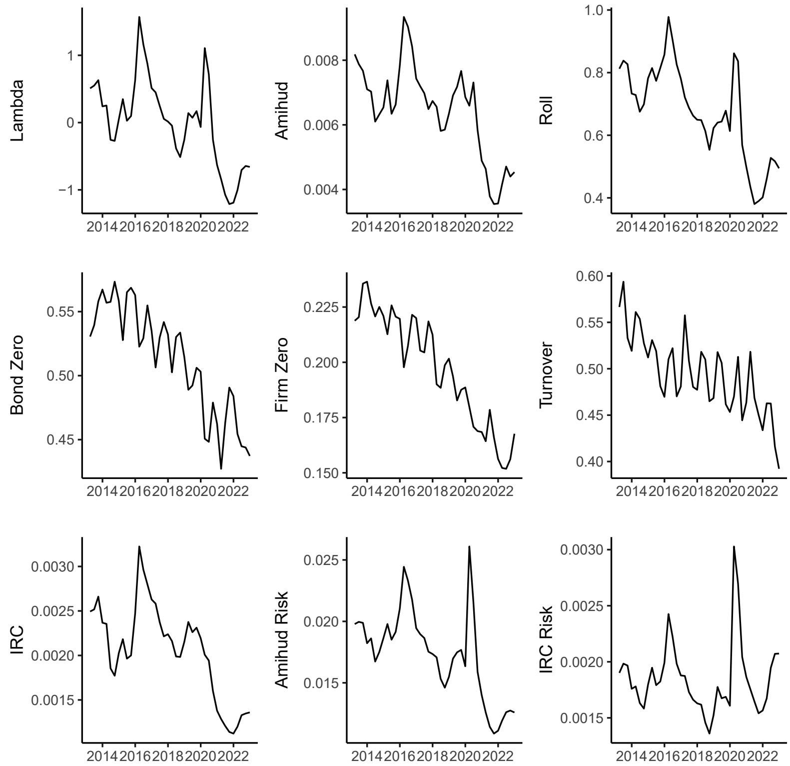
Panel A: Liquidity Percentiles

	λ	Amihud	Roll	Firm Zero	Bond Zero	TO	IRC	Amihud Risk	IRC Risk
99th	12.314	0.069	2.599	1.000	0.984	0.770	0.012	0.088	0.011
95th	6.346	0.025	1.667	1.000	0.953	0.425	0.006	0.053	0.006
75th	0.839	0.006	0.851	0.194	0.766	0.180	0.002	0.023	0.003
50th	-0.997	0.003	0.521	0.000	0.516	0.092	0.001	0.011	0.001
25th	-2.046	0.001	0.312	0.000	0.254	0.040	0.001	0.005	0.000
5th	-2.879	0.000	0.140	0.000	0.033	0.004	0.000	0.001	0.000
1st	-3.184	0.000	0.073	-0.016	0.000	0.000	0.000	0.000	0.000

Panel B: Correlation

	λ	Amihud	Roll	Firm Zero	Bond Zero	TO	IRC	Amihud Risk	IRC Risk
λ	1								
Amihud	0.768	1							
Roll	0.619	0.486	1						
Firm Zero	0.0002	0.008	0.048	1					
Bond Zero	-0.130	0.042	0.159	0.241	1				
TO	0.095	-0.070	-0.039	-0.071	-0.500	1			
IRC	0.853	0.565	0.509	0.052	0.020	0.064	1		
Amihud Risk	0.807	0.583	0.628	-0.015	-0.080	0.016	0.515	1	
IRC Risk	0.821	0.369	0.401	-0.041	-0.302	0.234	0.638	0.518	1

Figure 3: Time series of liquidity variables. This chart shows the time series of liquidity variables across our sample from January 2013 to December 2022 following Dick-Nielsen et al. (2012). Liquidity variables are calculated quarterly for each bond as explained in Appendix A. The cross-sectional mean is plotted for each liquidity variable in each quarter.



B Alternative Sampling

In this section, we present the results as outlined in Table 5 based on alternative samples. In Table B-1, we do not apply any restrictions on trade size, i.e., we also include retail-size trades. For Table B-2, we set the threshold for identifying retail-sized trades to $100,000 as it is common practice in empirical bond research. The results shown in Table B-3 are based on a subsample of exclusively non-financial companies, that is, excluding companies with a first-digit SIC code of 6. Finally, Table B-4 presents the results based on the 2004 time series. It should be noted that the sparse data in the earlier part of the sample translates into an average sample size of 47.74, which is much smaller compared to the later part of the sample, resulting in a much lower test power.

Eventually, Table B-5 to Table B-8 exhibit the test power conditional on bond liquidity as displayed in Table 11 and Table 12, but for alternative liquidity proxies, i.e., Roll and turnover.

Table B-1: Size and power tests including retail-sized trades. This table presents size and power tests of firm-level abnormal standardized returns (ABSR), standardized abnormal returns (SABR), unstandardized abnormal returns (ABR, raw), and ABSR with pre-event period $[t-101, t-6]$ standard deviation (ABSR-Pre). The tests include the parametric t-test (t), the non-parametric signed-rank test (SR), and the sign test (S). For each test, we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws. Panel A shows descriptive statistics of returns in percent. Panel B presents the test size using a two-sided significance level (SL) of 1%. Panel C shows the power tests for artificial positive and negative 15bp shocks. Panel D repeats the power tests for alternative shocks of 10bp and 25bp.

Panel A: Return Characteristics

	Mean	Median	SD	Skew	Kurt	N
ABSR	-0.0008	-0.0037	0.8563	0.0151	1.8889	5,138,467
SABR	-0.0056	-0.0085	1.0511	0.0265	1.4911	5,141,945

Panel B: Size Tests

	t		SR		S	
SL (%)	0.5	99.5	0.5	99.5	0.5	99.5
ABSR	0.24	0.55	0.32	0.44	0.34	0.26
SABR	0.28	0.53	0.42	0.47	0.49	0.30
ABR (raw)	0.19	0.26	0.35	0.57	0.39	0.31
ABSR-Pre	0.25	0.58	0.33	0.60	0.20	0.43

Panel C: Power Tests - 15bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
ABSR	94.05	98.88	98.27	95.31	99.33	98.59
SABR	96.02	99.05	98.22	96.20	99.27	98.02
ABR (raw)	29.27	87.74	94.66	31.78	88.85	94.87
ABSR-Pre	84.47	95.03	93.94	89.28	97.30	95.96

Panel D: Power Tests - 10bp and 25bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
10bp - raw	10.90	51.94	66.10	12.59	54.09	65.46
10bp - stand.	60.59	82.36	81.55	65.03	84.41	82.76
25bp - raw	69.13	99.84	99.97	71.39	99.91	100.00
25bp - stand.	100.00	100.00	100.00	99.99	100.00	100.00

Table B-2: Size and power tests excluding trades below $100,000. This table presents size and power tests of firm-level abnormal standardized returns (ABSR), standardized abnormal returns (SABR), unstandardized abnormal returns (ABR, raw), and ABSR with pre-event period $[t-101, t-6]$ standard deviation (ABSR-Pre). The tests include the parametric t-test (t), the non-parametric signed-rank test (SR), and the sign test (S). For each test, we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws. Panel A shows descriptive statistics of returns in percent. Panel B presents the test size using a two-sided significance level (SL) of 1%. Panel C shows the power tests for artificial positive and negative 15bp shocks. Panel D repeats the power tests for alternative shocks of 10bp and 25bp.

Panel A: Return Characteristics

	Mean	Median	SD	Skew	Kurt	N
ABSR	0.0000	-0.0031	0.7908	0.0370	2.6550	2,116,300
SABR	-0.0006	-0.0052	1.0761	0.0467	2.0869	2,118,803

Panel B: Size Tests

	t		SR		S	
SL (%)	0.5	99.5	0.5	99.5	0.5	99.5
ABSR	0.30	0.46	0.39	0.51	0.30	0.34
SABR	0.36	0.50	0.45	0.47	0.31	0.38
ABR (raw)	0.23	0.42	0.35	0.46	0.35	0.38
ABSR-Pre	0.27	0.36	0.42	0.59	0.35	0.42

Panel C: Power Tests - 15bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
ABSR	85.38	94.11	89.92	86.78	94.68	90.23
SABR	84.92	93.70	88.75	86.07	94.22	89.32
ABR (raw)	32.31	78.07	84.78	34.22	81.03	86.17
ABSR-Pre	65.90	81.51	75.00	69.80	85.14	78.46

Panel D: Power Tests - 10bp and 25bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
10bp - raw	12.41	42.18	48.76	14.07	45.01	50.63
10bp - stand.	48.40	67.17	60.97	51.74	68.88	61.22
25bp - raw	70.70	99.11	99.64	70.85	99.23	99.61
25bp - stand.	99.77	99.98	99.63	99.86	99.96	99.78

Table B-3: Size and power tests without financial companies. This table presents size and power tests of firm-level abnormal standardized returns (ABSR), standardized abnormal returns (SABR), unstandardized abnormal returns (ABR, raw), and ABSR with pre-event period $[t-101, t-6]$ standard deviation (ABSR-Pre). The tests include the parametric t-test (t), the non-parametric signed-rank test (SR), and the sign test (S). For each test, we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws. Panel A shows descriptive statistics of returns in percent. Panel B presents the test size using a two-sided significance level (SL) of 1%. Panel C shows the power tests for artificial positive and negative 15bp shocks. Panel D repeats the power tests for alternative shocks of 10bp and 25bp.

Panel A: Return Characteristics

	Mean	Median	SD	Skew	Kurt	N
ABSR	0.0004	-0.0032	0.7986	0.0490	2.4339	2,057,489
SABR	-0.0010	-0.0060	1.0658	0.0486	1.8251	2,059,299

Panel B: Size Tests

	t		SR		S	
SL (%)	0.5	99.5	0.5	99.5	0.5	99.5
ABSR	0.40	0.51	0.52	0.53	0.41	0.49
SABR	0.45	0.45	0.49	0.44	0.50	0.43
ABR (raw)	0.26	0.29	0.45	0.52	0.47	0.37
ABSR-Pre	0.27	0.43	0.35	0.52	0.32	0.42

Panel C: Power Tests - 15bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
ABSR	85.96	95.15	92.26	86.85	95.24	91.83
SABR	86.17	94.64	90.01	86.87	94.69	89.72
ABR (raw)	28.83	77.38	83.46	29.14	78.01	83.90
ABSR-Pre	68.62	84.53	79.24	71.15	86.27	80.73

Panel D: Power Tests - 10bp and 25bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
10bp - raw	11.10	40.27	47.34	11.67	40.97	45.91
10bp - stand.	48.44	67.95	61.90	49.55	67.85	61.10
25bp - raw	66.96	99.02	99.55	66.90	99.25	99.64
25bp - stand.	99.87	99.97	99.85	99.91	99.98	99.78

Table B-4: Size and power tests from 2004 to 2022. This table presents size and power tests of firm-level abnormal standardized returns (ABSR), standardized abnormal returns (SABR), unstandardized abnormal returns (ABR, raw), and ABSR with pre-event period $[t-101, t-6]$ standard deviation (ABSR-Pre). The tests include the parametric t-test (t), the non-parametric signed-rank test (SR), and the sign test (S). For each test, we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws. Panel A shows descriptive statistics of returns in percent. Panel B presents the test size using a two-sided significance level (SL) of 1%. Panel C shows the power tests for artificial positive and negative 15bp shocks. Panel D repeats the power tests for alternative shocks of 10bp and 25bp.

Panel A: Return Characteristics

	Mean	Median	SD	Skew	Kurt	N
ABSR	0.0005	-0.0022	0.8225	0.0191	2.1295	4,348,794
SABR	-0.0033	-0.0057	1.0726	0.0191	1.7107	4,354,046

Panel B: Size Tests

	t		SR		S	
SL (%)	0.5	99.5	0.5	99.5	0.5	99.5
ABSR	0.43	0.47	0.51	0.56	0.31	0.32
SABR	0.60	0.42	0.52	0.50	0.38	0.27
ABR (raw)	0.22	0.32	0.45	0.53	0.35	0.43
ABSR-Pre	0.29	0.48	0.37	0.56	0.21	0.34

Panel C: Power Tests - 15bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
ABSR	62.68	75.97	66.69	65.02	77.50	67.93
SABR	63.60	75.58	64.79	63.99	75.63	64.25
ABR (raw)	17.58	49.84	56.74	19.34	52.00	57.70
ABSR-Pre	45.77	60.70	51.60	49.47	64.29	54.63

Panel D: Power Tests - 10bp and 25bp Shocks

	Negative			Positive		
	t	SR	S	t	SR	S
10bp - raw	6.60	22.44	26.19	7.35	23.97	26.81
10bp - stand.	29.24	42.51	35.58	31.70	44.08	36.41
25bp - raw	48.81	88.07	91.35	51.31	89.69	92.35
25bp - stand.	96.04	97.90	94.13	96.48	98.15	94.69

Table B-5: Test power stratified by liquidity (Roll measure) and maturity for a 15bp shock. This table reports the test power of the signed-rank test at the 1% significance level based on bond-level abnormal standardized returns. We group bonds by five liquidity and four maturity cut-offs. The test power in Panel A is based on positive 15bp shocks (i.e., we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws). Panel B reports the corresponding number of observations.

Panel A: Power Test - 15bp Shocks

	1-3 years	3-5 years	5-10 years	> 10 years
High	100.00	100.00	100.00	36.37
2	100.00	100.00	99.97	92.93
3	100.00	100.00	99.91	93.92
4	100.00	99.99	99.53	87.72
Low	100.00	78.50	83.47	64.09

Panel B: Number of Observations

	1-3 years	3-5 years	5-10 years	> 10 years
High	281,698	163,440	176,533	29,355
2	172,274	164,459	229,758	86,404
3	108,522	157,601	249,769	135,208
4	61,149	139,485	281,793	166,208
High	37,218	88,610	286,949	231,846

Table B-6: Test power stratified by liquidity (Roll-measure) and rating for a 15bp shock. This table reports the test power of the signed-rank test at the 1% significance level based on bond-level abnormal standardized returns. We group bonds by five liquidity and six rating cut-offs. The test power in Panel A is based on positive 15bp shocks (i.e., we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws). Panel B reports the corresponding number of observations.

Panel A: Power Test - 15bp Shocks

	AAA to AA	A	BAA	BA	B	Below B
High	100.00	100.00	100.00	99.27	80.66	99.84
2	100.00	100.00	100.00	99.75	97.23	80.06
3	100.00	100.00	99.99	99.16	95.91	35.32
4	99.99	100.00	99.71	93.64	75.15	29.90
Low	98.18	98.19	81.70	69.73	43.58	10.41

Panel B: Number of Observations

	AAA to AA	A	BAA	BA	B	Below B
High	56,943	256,968	273,757	43,124	15,879	4,355
2	53,546	217,190	258,801	84,300	33,887	5,171
3	45,294	192,796	258,936	98,459	48,822	6,793
4	43,775	171,048	265,671	111,198	48,193	8,750
Low	30,358	106,744	308,721	123,055	63,694	12,051

Table B-7: Test power stratified by liquidity (turnover) and maturity for a 15bp shock. This table reports the test power of the signed-rank test at the 1% significance level based on bond-level abnormal standardized returns. We group bonds by five liquidity and four maturity cut-offs. The test power in Panel A is based on positive 15bp shocks (i.e., we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws). Panel B reports the corresponding number of observations.

Panel A: Power Test - 15bp Shocks

	1-3 years	3-5 years	5-10 years	> 10 years
HIgh	100.00	100.00	99.97	84.17
1	100.00	100.00	99.87	82.33
2	100.00	100.00	99.73	78.32
3	100.00	100.00	99.38	78.21
Low	100.00	99.99	98.87	78.93

Panel B: Number of Observations

	1-3 years	3-5 years	5-10 years	> 10 years
High	166,460	165,969	188,040	132,153
1	154,920	150,362	218,659	128,218
2	135,702	140,995	248,887	126,806
3	118,953	138,450	271,586	122,931
Low	86,245	119,973	302,089	143,833

Table B-8: Test power stratified by liquidity (turnover) and rating for a 15bp shock. This table reports the test power of the signed-rank test at the 1% significance level based on bond-level abnormal standardized returns. We group bonds by five liquidity and six rating cut-offs. The test power in Panel A is based on positive 15bp shocks (i.e., we show rejection rates (in percent) from 10,000 simulations of 300 firm×day draws). Panel B reports the corresponding number of observations.

Panel A: Power Test - 15bp Shocks

	AAA to AA	A	BAA	BA	B	Below B
High	100.00	100.00	100.00	45.13	4.06	0.18
1	100.00	100.00	100.00	92.77	74.12	7.11
2	100.00	100.00	100.00	97.19	91.72	23.15
3	100.00	100.00	99.99	98.89	89.07	26.16
Low	100.00	100.00	99.95	97.53	85.27	89.65

Panel B: Number of Observations

	AAA to AA	A	BAA	BA	B	Below B
High	71,029	269,843	261,926	35,550	12,176	2,098
1	59,056	220,328	273,596	67,029	28,591	3,559
2	46,519	178,203	278,990	97,570	45,641	5,467
3	33,841	154,077	270,859	127,323	57,836	7,984
Low	19,821	125,900	287,772	133,805	66,669	18,173