Essays on U.S. Corporate Bonds

D I S S E R T A T I O N
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by

Thomas Philipp Aeschbacher

from

Trachselwald (Bern)

Approved on the application of

Prof. Francesco Audrino, PhD

and

Prof. Dr. Axel Kind

Dissertation no. 5262

Difo-Druck GmbH, Untersiemau 2022

Essays on U.S. Corporate Bonds

D I S S E R T A T I O N
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by

Thomas Philipp Aeschbacher

from

Trachselwald (Bern)

Approved on the application of

Prof. Francesco Audrino, PhD

and

Prof. Dr. Axel Kind
Prof. Beatrix Eugster, PhD

Dissertation no. 5262

Difo-Druck GmbH, Untersiemau 2022

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, June 13, 2022

The President:

Prof. Dr. Bernhard Ehrenzeller

I thank my supervisor Francesco Audrino, my parents, and my friends for their constant support throughout this journey.

Summary

In the first paper, I analyze fallen angel bonds' returns before and after their downgrade to high-yield. Fallen angel bonds experience a sharp price decline prior and a sharp recovery after the rerating announcement by the rating agency. I introduce a novel benchmark that should more closely mirror the price fallen angel bonds would have had, had they not experienced a fire sale prior to their downgrade. This allows me to estimate the total preannouncement sell-off and I then use this in order to decide on which fallen angel bonds should be bought after their downgrade. There exists a strong negative relationship between the size of the preannouncement sell-off and future postannouncement returns. I can then show, that an investor fares better, if only fallen angels are bought that trade at a discount to their respective estimated benchmark return had they not experienced the sell-off. Using the introduced innovative, data-driven, novel benchmark allows an investor to generate higher returns. This outperformance gets more pronounced as one focuses on the fallen angel bonds that experienced the highest preannouncement sell-off.

In the second paper, we study the cross-section of corporate bonds utilizing a large set of financial statements, equity and bond characteristics. We use a predictive regression framework and the adaptive Lasso to choose the most relevant characteristics for the cross-section of corporate bonds. Applying the adaptive Lasso to the full dataset, we find a ten-factor model, with value, bond reversal, and equity momentum spillover being the dominant factors. Contrary to equity studies, financial variables from Compustat do not appear to have strong power in predicting corporate bond returns. We validate our initial results by running an out-of-sample exercise using an expanding window approach. Out of the 60 months utilized in the out-of-sample, the adaptive Lasso consistently chooses value, bond reversal, and equity momentum spillover. Finally, we evaluate the economic benefits of investing according to the predictions of the adaptive Lasso and find significant benefits in terms of absolute and risk-adjusted returns.

In the third paper, we evaluate the ability of U.S. corporate bond fund managers to generate alpha. We apply the False Discovery Rate (FDR) to distinguish between "skill" and "luck." We find that long-term outperformance remains elusive, with only 1% of the funds able to generate significant alpha over their life. However, fund managers are able to generate alpha over the short-term with the proportion of skilled funds increasing to 13.5% when we examine three-year sub-periods. To confirm these findings, we design an out-of-sample investment strategy where we invest in funds according to their estimated "skill" from past returns. Our strategy generates positive and significant alpha, which confirms the persistence in outperformance over the short-run. Our results are economically meaningful for investors suggesting that dynamic and active manager selection pays off.

Zusammenfassung

In der ersten Arbeit analysiere ich die Renditen von “Fallen Angel”-Anleihen vor und nach ihrer Herabstufung zu hochverzinslichen Anleihen. “Fallen Angel”-Anleihen verzeichnen einen starken Kursrückgang vor und eine deutliche Erholung nach der Ankündigung der Herabstufung durch die Ratingagentur. Ich führe eine neue Benchmark ein, die den Preis von “Fallen Angel”-Anleihen besser widerspiegeln sollte, wenn sie nicht vor ihrer Herabstufung einen Ausverkauf erlebt hätten. Auf diese Weise kann ich den gesamten Ausverkauf vor der Bekanntgabe der Herabstufung schätzen und auf dieser Grundlage entscheiden, welche “Fallen Angel”-Anleihen nach ihrer Herabstufung gekauft werden sollten. Es besteht ein starker negativer Zusammenhang zwischen dem Ausmass des Ausverkaufs vor der Herabstufung und den zukünftigen Renditen nach der Herabstufung. Ich kann dann zeigen, dass ein Investor höhere Renditen erzielen kann, wenn er nur “Fallen Angel”-Anleihen kauft, die mit einem Abschlag zu ihrer jeweiligen geschätzten Benchmark-Rendite gehandelt werden. Im Vergleich zu bisherigen Methoden aus der Literatur, kann ein Anleger mit der eingeführten innovativen, datengestützten, neuartigen Benchmark höhere Renditen erzielen. Diese Outperformance wird umso deutlicher, je mehr man sich auf die “Fallen Angel”-Anleihen konzentriert, die vor der Ankündigung den grössten Ausverkauf erlebt haben.

In der zweiten Arbeit untersuchen wir den Querschnitt von Unternehmensanleihen, indem wir einen grossen Datensatz von Jahresabschluss-, Aktien- und Anleihenmerkmalen verwenden. Wir verwenden einen prädiktiven Regressionsrahmen und das adaptive Lasso, um die wichtigsten Merkmale für den Querschnitt der Unternehmensanleihen auszuwählen. Bei Anwendung des adaptiven Lasso auf den gesamten Datensatz ergibt sich ein Zehn-Faktoren-Modell, wobei Value, Bond Reversal und Equity Momentum Spillover die dominierenden Faktoren sind. Im Gegensatz zu Aktienstudien scheinen Finanzvariablen aus Compustat keine grosse Aussagekraft bei der Vorhersage von Unternehmensanleihenrenditen zu haben. Wir validieren unsere ersten Ergebnisse, indem wir eine Out-of-Sample-Untersuchung mit einem Expanding-Window-Ansatz durchführen. Von den 60 Monaten, die in der Out-of-Sample-Stichprobe verwendet wurden, wählt das adaptive Lasso konsistent Value-, Bond-Reversal- und Equity-Momentum-Spillover aus. Abschliessend bewerten wir den wirtschaftlichen Nutzen von Investitionen gemäss den Vorhersagen des adaptiven Lassos und stellen fest, dass sie in Bezug auf die absoluten und risikobereinigten Renditen erhebliche Vorteile bieten.

In der dritten Arbeit wird die Fähigkeit der Manager von Unternehmensanleihenfonds, Alpha zu generieren, bewertet. Wir wenden die False Discovery Rate (FDR) an, um zwischen “Können” und “Glück” zu unterscheiden. Wir stellen fest, dass eine langfristige Outperformance schwer zu erreichen ist, da nur 1% der Fonds in der Lage ist, während ihrer Laufzeit ein signifikantes Alpha zu erzielen. Allerdings sind die Fondsmanager in der Lage, kurzfristig Alpha zu generieren, wobei der Anteil der “Können” Fonds auf 13,5% steigt, wenn wir dreijährige Teilperioden untersuchen. Um diese Ergebnisse zu validieren, entwerfen wir eine Out-of-Sample-Anlagestrategie, bei der wir in Fonds entsprechend ihrem geschätzten “Können” investieren. Unsere Strategie generiert ein positives und signifikantes Alpha, was die Persistenz der Outperformance im kurzfristigen Bereich bestätigt. Unsere Ergebnisse sind für Anleger von wirtschaftlicher Bedeutung und deuten darauf hin, dass sich eine dynamische und aktive Managerauswahl auszahlt.

Contents

Chapter 1

The preannouncement sell-off of fallen angel bonds: An event study using a novel benchmark

Abstract

In this paper, I analyze fallen angel bonds' returns before and after their downgrade to high-yield. Fallen angel bonds experience a sharp price decline prior and a sharp recovery after the rerating announcement by the rating agency. I introduce a novel benchmark that should more closely mirror the price fallen angel bonds would have had, had they not experienced a fire sale prior to their downgrade. This allows me to estimate the total preannouncement sell-off and I then use this in order to decide on which fallen angel bonds should be bought after their downgrade. There exists a strong negative relationship between the size of the preannouncement sell-off and future postannouncement returns. I can then show, that an investor fares better, if only fallen angels are bought that trade at a discount to their respective estimated benchmark return had they not experienced the sell-off. Using the introduced innovative, data-driven, novel benchmark allows an investor to generate higher returns. This outperformance gets more pronounced as one focuses on the fallen angel bonds that experienced the highest preannouncement sell-off.

1.1 Introduction

This paper studies the behavior of corporate bond prices prior and post to their downgrade to high-yield. When a corporate bond is downgraded from investment-grade to high-yield, this is referred to a fallen angel event. Why do investment professionals refer to this as a fallen angel event? To cite Clare, Thomas, and Motson (2016), one could think of fallen angels to represent angels that have been rejected from heaven due to their unbecoming behavior. Their unbecoming behavior is that they once had an investment-grade rating from Moody's or Standard & Poor's (S&P), but were downgraded and no longer possess an investment-grade rating from either (Ambrose, Cai, & Helwege, 2008). This means that the once investment-grade rated corporate bond got forced into the high-yield fixed income asset class. This happens when the credit rating agency perceives a decline in the credit quality of the issuer high enough to justify a speculative-grade rating. Emery and Gates (2014) examines the press releases from Moody's and concludes that 72 percent are due to company specific reasons and industry stress. Why should a downgrade to high-yield be bad? High-yield bonds tend to be riskier because their risk of default is higher. Therefore many investment and fund managers must sell them from their portfolios because they are constrained by their investment guidelines. Investment guidelines specify the risks that are allowed to be taken in the portfolio and refer specifically to ratings from Moody's or S&P. Furthermore, passive investment vehicles – such as exchange traded funds – have gained a lot of traction in recent years. Because exchange traded funds aim to replicate their specific investment-grade benchmark index, they naturally sell-off the fallen angel bonds once they get downgraded to high-yield. In addition, investment professionals tend to anticipate which bonds are likely to get downgraded given that public information on bond issuers is updated regularly and sell them well in advance of the downgrade announcement by the rating agency. This is because fund managers do not like to report downgrades in their investment-grade rated fund.

The literature on bond rating changes and associated bond price effects, documents for fallen angels following stylized facts: They tend to underperform investment-grade bonds before the announcement and outperform high-yield bonds after the downgrade. This implies a "V-shaped" curve of fallen bond prices, centered at their downgrade date.

As already outlined, this particular anomaly is likely to have its roots in institutional and behavioral factors. Investment managers must adhere to strict investment guidelines and there also seems to be evidence of fund managers engaging in a precautionary sell-off. This results in fallen angel bonds experiencing a sell-off prior to their downgrade to high-yield. This sell-off likely pushes the price of fallen angel bonds below their "fair value." The key question hereby is, what is a fallen angel's respective fair value? Or stated otherwise: What would the price of the bond have been, had it not been sold-off in a fire sale? In econometrics, this hypothetical, unobservable outcome, is known as a counterfactual.

In prior literature, the fair value is simply calculated as a difference between the returns of the fallen angel bond and a broad bond benchmark index. These differences are added up in an event window around the downgrade announcement and averaged by firm. The methodology outlined is known as an event study.

Estimation of these effects could be problematic with the existing event study literature, in part because of the choice of the broad bond benchmark used as a counterfactual. This is because at the time of price movements related to a potential downgrade, fallen angel bonds differ from the broad investment-grade market with respect to individual characteristics that arguably relate to the probability of being downgraded and becoming subsequently a fallen angel bond. A simple difference of the time-series return of fallen angels and investment-grade bonds would therefore not only reflect the effect of becoming a fallen angel bond but also the pre-anticipation period differences in credit and issuer quality that affect the future price path of the return on corporate bonds.

I am interested in estimating the preannouncement sell-off fallen angel bonds suffer. Furthermore, I then aim to investigate whether the difference between the fallen angels' bond price and its estimated fair value at the date of the downgrade can be used to explain future fallen angel bond returns. This research question is relevant for many different reasons. Firstly, it shows how big the sell-off due to institutional and behavioral factors might be on average. Secondly, if there is a relation between the size of the preannouncement sell-off and future bond returns, rich investment opportunities could arise. Fallen angel factor portfolios could be constructed, targeted at specifically harvesting this anomaly. Questions such as how long to keep the position in the portfolio arise naturally and practical guidance can be given with this study. Furthermore, it presents investment-grade corporate bond portfolio managers with an incentive to strategically drift from their benchmark. If they sell-off the asset at the same time everyone else does, they simply lock-in their losses. If they, however, strategically time their exit – which they have to do, due to restrictions on credit quality in their investment guidelines – they could recoup some of their loss. This study could therefore also impact the literature in the creation of "better" investment guidelines for investment professionals, giving them more leeway in the time that they have to sell the asset that has been downgraded to high-yield.

In this study, I try to merge the traditional event study framework with a framework from the causal inference literature that is specifically targeted at estimating these kind of effects under a certain set of identifying assumptions. Thus, the unified framework outlined in this paper could also be applied to similar event studies that try to find a better, more "custom" benchmark. The core of the causal analysis revolves around the imputation of the unknown, unobservable counterfactual of the fallen angel bond. Different methodologies exist to estimate this unknown counterfactual, whereas the estimation in general is concerned with estimating a vector of weights for the control units that minimizes the distance in pre-treatment outcomes between the treated unit (the fallen angel bond) and pool of controls (investment-grade bonds that did not experience a downgrade to high-yield). Pre-treatment refers to the period before fund managers and investment professionals engage in a sell-off because they anticipate a downgrade. A popular and natural estimation procedure for this type of problem is the synthetic control method by Abadie, Diamond, and Hainmueller (2010). The procedure selects a set of weights such that chosen covariates and pre-treatment outcomes of the treated unit are approximately matched by a weighted average of the control units. Abadie et al. (2010) impose a no-intercept condition as well as that the weights need to sum up to one and non-negativity of the weights. However, as noted by Athey and Imbens (2017), it is not obvious that restricting the weights to be positive and sum up to one is always the most sensible choice for constructing the counterfactual. Considering this and the fact that my particular estimation problem is of a high-dimensional nature, i.e. the number of controls is larger than the sample size, I use an approach in the flavor of Doudchenko and Imbens (2017), who propose to use an elastic-net type penalty for the weights and do not impose any of the aforementioned restrictions on the weights or intercept. Popular other contestants of this estimation problem are difference-in-differences (DID), constrained regression and best subset selection. Similarly like the synthetic control method by Abadie et al. (2010), these methods also frequently impose next to a no-intercept condition a collection of constraints on the weights, e.g. weights have to add up to one, non-negativity or even constant-weights in the case of DID.

In the empirical part, I use a comprehensive dataset and am able to show the notorious "V-shape" curve documented in the literature. The results show, that during the preannouncement period, fallen angel bonds suffer a substantial sell-off. Furthermore, after the sell-off has settled, fallen angel bonds recoup the bulk of the value they lost due to the downgrade to high-yield and thus offer substantial investment opportunities. The size of the preannouncement sell-off is negatively related to the future return of fallen angel bonds. In the data, there seems to be an indication, that buying fallen angel bonds that trade below their estimated fair value outperforms a naive strategy that disregards information about their preannouncement sell-off. A holding period of 10 months reaps the bulk of the outperformance with respect to the new broad high-yield bond benchmark index. Using my "novel" benchmark that was estimated with the synthetic control, I can show that – albeit its outperformance not being statistically significant – it offers higher returns than when applying other choices of benchmarks often encountered in the literature.

The remainder of the paper is organized as follows. Section 1.2 provides a brief overview of the main studies that analyze the impact of rating changes on corporate bond prices. Section 1.3 gives a review of traditional event studies and section 1.4 introduces the causal event study framework. Section 1.5 presents the data and section 1.6 reports the estimation results. Section 1.7 concludes.

1.2 Literature on bond rating effects

A vast body of literature investigates changes in credit ratings on corporate bond prices. The majority of these studies are conducted with the event study methodology and find significant negative average excess returns associated with downgrades. In event studies, time-series of affected units are centered around an event and excess returns are taken with respect to some unaffected benchmark. These "abnormal returns" are then added up which yields "cumulative abnormal returns." Cumulative abnormal returns are then grouped on the firm level over which the average is taken. The literature on bond rating effects uses as an event naturally the rerating announcement of the rating agency and for the unaffected benchmark a simple market-value weighted bond benchmark. In the United States (U.S.) there exist three big rating agencies: Moody's, S&P and Fitch. Studies concerned with U.S. corporate bonds predominantly utilize data from Moody's and S&P, since Fitch has a smaller market share and therefore fewer ratings on corporate bonds. May (2010) collects data from all three rating agencies and the tabulations show that Moody's and S&P with a share of 42 and 40 percent, respectively, contribute far more towards the final sample than Fitch with a share of only 17 percent. In the past, the literature utilized predominantly monthly or weekly data, but has now shifted to daily data due to the larger availability of high-quality data sources and research showing that monthly returns lack the power to detect excess returns (Bessembinder, Kahle, Maxwell, & Xu, 2009). Concerning the results there is substantial variation with regard to the time period relative to the downgrade event where significant effects are found. Furthermore, not all studies investigate the more narrow question of price effects of a bond becoming a fallen angel. A large part of the literature investigates incremental changes in ratings without specifically analyzing the investment-grade versus high-yield threshold. This can in part be explained by the relative importance of fallen angel bonds over time. Whereas in the period 1997 to 2002 only 10–15 percent of all outstanding high-yield corporate debt corresponded to fallen angels, this share has increased to 25 percent in the year 2009 (Bolognesi, Ferro, & Zuccheri, 2014).

One of the earliest research papers on this topic was written by Grier and Katz (1976). Grier and Katz (1976) investigate the impact of reclassification in ratings on bond prices from industrial and utility bonds. They find that there is an anticipation period in the industrial bond market to rating changes and that their reaction relative to utility bonds is more pronounced. Furthermore, Grier and Katz (1976) find that the negative price effect occurs in the month of and the one following the rating announcement. In their study they utilize rating changes from 1966 to 1972 and monthly data. Contrasting the results of Grier and Katz (1976) is the study of Weinstein (1977). Weinstein (1977) also analyzes the price behavior of corporate bonds during the period surrounding a rating change announcement. The author finds evidence of a price change ranging from 18 to 7 months prior to the announcement of the rating change. However, Weinstein (1977) does not find any evidence during the 6 months prior to the rating change, nor any during the month of the change or for 6 months postannouncement. Similarly, Wansley and Clauretie (1985) also use monthly returns and do not find any significant reactions to downgrades or upgrades in the month of and month following a rerating announcement. Hite and Warga (1997) analyze the effect of rating changes for industrial firms from 1985 to 1995 using monthly data.

They find strong negative effects during the event month that last up to 6 months after the rerating announcement for downgrades remaining below the investment-grade threshold. Furthermore, Hite and Warga (1997) are able to show that the downgrade effect is substantially higher for high-yield rated than for investment-grade firms. Wansley, Glascock, and Clauretie (1992) study the impact of bond reratings by S&P on bond prices using weekly data from 1982 to 1984. In the case of downgrades, they find a strong negative announcement effect in the week of the downgrade announcement. Furthermore, they find negative responses up until 3 weeks prior to the downgrade. Using daily data from 1977 to 1982 and bond downgrade data from S&P and Moody's, Hand, Holthausen, and Leftwich (1992) are able to further corroborate the negative average bond returns around the announcement day. In addition, Hand et al. (1992) show that their announcement effect is stronger for high-yield bonds than for investment-grade rated bonds. Steiner and Heinke (2001) focus on the German Eurobond market and consider rating changes from 1985 to 1996. They use daily returns for their analysis and find a strong reaction to downgrades on the announcement day, especially for downgrades into speculative-grade. Furthermore, Steiner and Heinke (2001) find significant price movements up to 100 days prior to the rerating announcement. In addition to the downwards trend prior to the downgrade, they also detect positive excess returns after the downgrade announcement between day 15 and 45. Steiner and Heinke (2001) conclude, that in particular for fallen angel bonds the negative announcement effects can in part be explained by price pressure effects due to regulatory constraints rather than original information content of rating changes. Furthermore, they explain their finding of a rebound in the postannouncement period with investors' attitude to overreact to downgrades. May (2010) analyzes rating changes during the period 2002 to 2009 with daily data and finds evidence of negative excess returns in the month before the downgrade announcement. The author's estimated negative effect immediately following the rerating announcement as well as up until 10 days postannouncement is highly significant. Dor and Xu (2011) corroborate the existence of a significant three month negative preannouncement trend in excess returns. Their negative price effect is the strongest during the month of the rerating event. Finally, Bolognesi et al. (2014) study the impact of a formerly investment-grade rated European corporate bond being downgraded to high-yield status. Their analysis therefore is one of the few papers that exclusively covers fallen angel bonds. In their analysis they cover the period from 2001 to 2009 and rating changes from Moody's and S&P. The authors construct their set of fallen angel bonds by monitoring a specific Merrill Lynch investment-grade index. According to the index rules a fallen angel event is when both Moody's and S&P downgrade the bond to speculative-grade. The results of Bolognesi et al. (2014) are threefold: Firstly, the first downgrade event as well as the second downgrade feature statistically significant and large negative preannouncement returns, both for the (-30, -1) and (-15, -1) days event window. In addition to the preannouncement windows, the same results hold both in terms of significance and magnitude for the (-1, 1) event window covering the immediate postannouncement period. Secondly, their results do not differ significantly for the first and the second downgrade announcement by the respective rating agencies, suggesting that the announcements happen not too far apart and that the respective windows are heavily correlated. Thirdly, Bolognesi et al. (2014) study the impact of deletion from the index which occurs at each month-end index reweighting date. They are able to show that post-deletion there exist significant positive excess returns, suggesting that it can be advantageous for investment managers to strategically time the buying of fallen angel bonds and drift from their benchmark. Furthermore, their cross-sectional results show that the bond's price rebound is proportional to the strength of the impact of the downgrade to fallen angel. Bolognesi et al. (2014) conclude that the observed outperformance of fallen angel bonds after the index rebalancing with respect to their high-yield peers relates to portfolio governance rules of many institutional investors. In particular, their ban on holding high-yield, non-investment-grade rated securities and the subsequent obligation to sell them once they reach fallen angel status and get deleted from the investment-grade index.

Regarding the mixed results from less contemporary literature, e.g. Weinstein (1977) and Wansley et al. (1992), there are two important key points regarding the quality of data that must be noted and that directly affect the validity of their obtained results. Firstly, the use of data on bond trades from the New York Stock Exchange which is characterized by infrequent trading and constitutes for a negligible fraction of overall market size and activity (May, 2010). Furthermore, the use of monthly returns and the resulting weaker power of tests, especially in small samples, is known and shown in e.g. Brown and Warner (1985) and Bessembinder et al. (2009). Secondly, the use of so called "matrix prices." Matrix prices are bond price estimates provided by commercial bond pricing services such as e.g. the Merrill Lynch Bond Pricing Service. The concern thereby being that matrix prices are predicted and not actual and ratings are one of the predictors. In case of downgrades, there would therefore be a mechanical strong negative price effect (May, 2010). It is therefore considered best practice when conducting studies on the U.S. corporate bond market to use daily data (or intra-day data aggregated to daily data as established by Bessembinder et al. (2009)) from the over-the-counter (OTC) market.

The literature review suggests a tentative pattern with respect to the price behavior of fallen angel bonds around their downgrade announcement. Namely that prior to the downgrade announcement there are significant negative cumulative abnormal returns, i.e. fallen angel bonds lose relative to their peers. The day of the announcement and the following day, bond prices tend to exhibit a significant decline. This decline tends to last a couple of days before the fallen angel bonds rebound again. The resulting pattern of cumulative abnormal returns therefore seems to represent a "V-shape" curve.

The causes of this particular phenomenon are likely to represent the institutional character of the U.S. corporate bond market. The reason for this is that corporate bonds trade OTC and that individual block sizes are too big for retail investors. Trading in corporate bonds is predominantly done by investment funds, passive investment vehicles and mutual funds (Bolognesi et al., 2014). However, investment funds such as insurance funds are subject to rigorous regulatory constraints based on credit rating. This means that these regulatory constraints can either restrict or forbid the ownership of non-investment-grade securities to retain the risk capital for other risky assets. Furthermore, investment funds also face reputational considerations (Ambrose et al., 2008; Ellul, Jotikasthira, & Lundblad, 2011). This means that investment funds that are constrained by their investment guidelines must sell downgraded securities immediately without considering current market prices or opportunities. Ambrose, Cai, and Helwege (2012) therefore also argue that insurers sell fallen angel bonds quicker in response to regulatory pressure. Cantor, Gwilym, and Thomas (2007) surveyed 200 U.S. and European fixed income fund managers and plan sponsors. The results of their study corroborate the importance of ratings in investment guidelines. They find that only 14 percent of the fund managers and 8 percent of the plan sponsors do not specify any explicit reference to ratings in their investment guidelines. The story with passive investment funds such as exchange traded funds that track an investment-grade bond index is similar. Since their objective is to replicate their respective benchmark, they will sell-off assets immediately that lose their investment-grade status and that get removed from the index. This particular setting is also analyzed by Bolognesi et al. (2014) who find that after the index finishes its rebalancing and removes the fallen angel bonds, they feature significant positive abnormal returns with respect to their new high-yield peers. Institutional factors are closely linked to insights from behavioral finance. Specifically, the tendency of individuals to overreact to bad news and underreact to good news. This is known as the overreaction hypothesis (Kahneman & Tversky, 1982). In a seminal study De Bondt and Thaler (1985) illustrated the overreaction hypothesis using U.S. stock data. Their finding was that stocks that had experienced significant losses tended to rebound in a predictable way over subsequent months, outperforming stocks that fared well initially. The conclusion is that investment managers are likely to sell-off securities well in advance of a potential downgrade, because they do not want to – or are not allowed to – report a downgrade of their portfolio. This leads to a "fire sale" that is naturally bound to be followed by a rebound.

1.3 Traditional event study framework

The literature on changes in credit ratings and their effect on corporate bond prices relies heavily on the event study methodology. In event studies there is an event some units are exposed to and the goal is to analyze the impact of the event on the treated units, i.e how their excess returns relative to peers behave around the event. To conduct an event study it is therefore necessary to identify the date of the event, the affected units and some form of control group. Time-series for each treated and its corresponding control group are shifted such that day 0, i.e. the event, is in the middle of the event window. In the case of rating changes and their impact on bond prices the event is given by the date of the downgrade announcement by the rating agency. The treated unit is the fallen angel bond and the controls are some form of corporate bond benchmark index. Subsequently, abnormal returns are formed. Daily abnormal bond returns are calculated as the raw return minus the contemporaneous return on an index of matched corporate bonds (May, 2010):

$$ABR_t = R_t - IR_t, \tag{1.1}$$

where ABR_t denotes the abnormal bond return on day t, R_t the raw bond return and IR_t the return on an index of matched corporate bonds that did not experience a downgrade to speculative-grade during the event window. Bessembinder et al. (2009) defines the raw bond return R_t as following:

$$R_t = \frac{P_t - P_{t-1} + AI_t}{P_{t-1}}, \tag{1.2}$$

where P_{t-1} and P_t are the daily prices on days $t-1$ and t, respectively. AI_t corresponds to the interest accrued over day t and is calculated as the annual coupon payment multiplied by L, all divided by 360, whereby L corresponds to the number of calendar days elapsed between the close of market day $t-1$ and the close of day t (May, 2010). In the literature and in common data provider services the raw bond return R_t is also known as the total return of a bond. Afterwards, the cumulative abnormal return (CAR) for each issue i over a test window from day H to day L is formed by simply summing up their respective abnormal returns:

$$CAR_{(H,L),i} = \sum_{t=H}^{L} ABR_{i,t}. \tag{1.3}$$

In some instances multiple issues i from the same firm j are downgraded at the same date, therefore cumulative abnormal returns are aggregated by firm to avoid positive correlation from bonds issued from the same firm. In the literature two common ways are adopted, either by equal-weighting the cumulative abnormal returns of the individual issues of each firm (Grier & Katz, 1976; Bolognesi et al., 2014), or weighting them by their respective share of the overall market-value of debt outstanding for each firm (Bessembinder et al., 2009):

$$CAR_{(H,L),j} = \sum_{i=1}^{n} CAR_{(H,L),i} w_{i,t-1}, \tag{1.4}$$

where $w_{i,t-1}$ represents in case of the equal-weighting aggregation scheme a constant of the size $1/n$, n indicating the number of issues i for each firm j. For the market-value weighting method, $w_{i,t-1}$ is the ratio of issue i's market value on day $t-1$ to the total market value of the firm j's issues. The cumulative abnormal returns are usually calculated in several pre-event and post-event windows to detect dynamics around the event. Inference is conducted by analyzing the vectors of $CAR_{(H,L),j}$ and performing classical t-tests as well as non-parametric Wilcoxon signed-rank tests against the null hypothesis that the mean of $CAR_{(H,L)}$ equals zero.

So far, so good. But what about IR_t? IR_t represents a return on an index of matched corporate bonds that did not experience a rating change. Although it is not always explicitly stated, what these studies analyzing the impact of rating changes on corporate bonds prices want to study is the causal effect. IR_t therefore represents the classical counterfactual, unobservable outcome from the causal inference literature. In contemporary and past literature only crude approximations based on ad hoc decisions of this counterfactual are utilized. May (2010) – who does not focus exclusively on fallen angel bonds

– constructs IR_t using 13 bond indices, resulting from a partitioning scheme that considers both the rating and the maturity of bonds. The maturity cutoff, i.e. long-maturity vs. short-maturity bonds, for 6 of the 7 rating categories are chosen such that there exist more or less an equal amount of bonds in given bond index. Bolognesi et al. (2014) consider in their study exclusively fallen angel bonds. In the pre-event period, i.e. before the bond gets downgraded, Bolognesi et al. (2014) simply take the broad market-value weighted investment-grade corporate bond index to measure the counterfactual. The counterfactual measures the unobservable return of what would the return of the fallen angel bond have been, had it not been downgraded. Is it really sensible to compare a bond that has suffered a decline in the perceived ability to repay its debt with the broad investment-grade market? Fallen angel bonds are usually at the lower end of the investment-grade credit spectrum already and might suffer from some kind of deterioration in financial metrics well in advance of a downgrade. They might already have suffered some downgrades within the investment-grade credit spectrum and suffered associated price shocks. Furthermore, the bulk of investment-grade debt tends to be offered by large corporations which tend to fare better in terms of credit rating, thus pushing up the average credit quality of a market-value weighted broad investment-grade bond index. In case one can not fully support the statement that the broad investment-grade market serves as a sensible counterfactual for fallen angel bonds, the estimated cumulative abnormal returns are not causal in nature. They simply document some form of excess return over broad investment-grade or high-yield investments around the event window. Rather, identifying assumptions must be made and an unknown, unobservable counterfactual – i.e. what would the price of the fallen angel bond have been, had it not been downgraded – has to be estimated.

In the next section I introduce a unifying framework for causal event studies. I propose to estimate IR_t, i.e. the return on an index of matched corporate bonds that did not experience a downgrade to speculative-grade, using a generalized variant of the synthetic control method.

1.4 Causal event study framework

Event studies have a very specific setup. In particular, they can be framed such that there exists one treated unit with an associated pool of potential control units. In the case of analyzing the effect of a downgrade from investment-grade to speculative-grade on bond prices, the treated unit is obviously the fallen angel bond. The pool of controls consist of all other investment-grade rated bonds that did not experience a downgrade to speculative-grade. At some point, there is a shift in perceived perception about the ability of the issuer of the corporate bond to be able to repay the debt. Similarly, market participants might expect the issue to be downgraded and thus want to avoid this by selling the bond. This is when things start to drift off for fallen angel bonds. This can be argued to be similar to e.g. the implementation of a new policy or some kind of treatment often studied in comparative case studies.

To cope with these kind of econometric problems, Abadie et al. (2010) introduced in a seminal paper the synthetic control method, see also Abadie, Diamond, and Hainmueller (2015) and Abadie and Gardeazabal (2003). The synthetic control method of Abadie et al. (2010) is designed to estimate the effect of a treatment in the presence of a single treated unit and a pool of control units. Their proposed method selects a set of weights such that chosen covariates and pre-treatment outcomes of a treated unit are approximately matched by a weighted average of control units. This weighted average of control units is coined synthetic control by Abadie et al. (2010). The synthetic control is then in turn used to estimate the unobserved counterfactual post-treatment in order to assess the impact of the treatment. The estimated causal effect of the treatment is then simply the difference between the observed time-series of the treated unit and the predicted time series using the synthetic control at each point in time where the policy takes place. The impact of the synthetic control method by Abadie et al. (2010) has been huge in the policy analysis literature, with a vast body of literature generalizing assumptions and working on improving inference methods. Athey and Imbens (2017) even state that the synthetic control method is one of the most important innovations in the program evaluation literature of the past fifteen years.

Doudchenko and Imbens (2017) generalize the synthetic control method and provide a unified framework for various other reduced-form approaches. Reduced-form approaches are characterized by modeling the counterfactual directly as a linear combination of the control units.

1.4.1 Framework

The general notation and framework follows the setting of Doudchenko and Imbens (2017). For ease of exposition I cover the case of a single treated unit, unit 0, and an associated single event. I consider a panel data setting with $N + 1$ cross-sectional units observed over the full sample in time periods $t =$

$1, \dots, T$. From period T_0+1 onwards, for $1, \dots, T_0, T_0+1, \dots, T$, the treated unit receives the treatment. This means that there are T_0 pre-treatment periods. Utilizing the potential outcome setup developed by Rubin (1974), there exists for each treated unit in each of the time periods, $t = T_0+1, \dots, T$ a pair of potential outcomes $Y_{0,t}(0)$ and $Y_{0,t}(1)$. $Y_{0,t}(0)$ corresponds to the potential outcome of the treated unit 0 at time t, given that the treated unit had not received the treatment. The treated unit actually received the treatment and therefore $Y_{0,t}(0)$ represents the counterfactual and is therefore unobserved. $Y_{0,t}(1)$ denotes the potential outcome of the treated unit 0 at time t given that the treated unit received the treatment. This potential outcome is therefore observed. The causal effects $\tau_{0,t}$ for the treated unit 0 for each time period t are as follows:

$$\tau_{0,t} = Y_{0,t}(1) - Y_{0,t}(0), \quad \text{for } t = T_0+1, \dots, T. \tag{1.5}$$

Units $i = 1, \dots, N$ are control units which do not receive the treatment in any of the time periods $t = 1, \dots, T$. In case of only wanting to estimate average treatment effects on the treated, the only potential outcome of interest corresponds to $Y_{i,t}(0)$ and is observed. To formalize the observed outcome a treatment indicator variable $W_{i,t}$ is necessary:

$$W_{i,t} = \begin{cases} 1 & \text{if } i = 0, \text{ and } t \in \{T_0+1, \dots, T\} \\ 0 & \text{otherwise} \end{cases} \tag{1.6}$$

For unit i in period t, the treatment $W_{i,t}$ and the realized outcome $Y_{i,t}^{\text{obs}}$ is observed:

$$Y_{i,t}^{\text{obs}} = Y_{i,t}(W_{i,t}) = \begin{cases} Y_{i,t}(0) & \text{if } W_{i,t} = 0 \\ Y_{i,t}(1) & \text{if } W_{i,t} = 1 \end{cases} \tag{1.7}$$

We can denote $\mathbf{Y}_{\text{i}}^{\text{obs}}$ as a $T \times 1$ vector $(Y_{i,1}^{\text{obs}}, \dots, Y_{i,T}^{\text{obs}})^T$. This vector we can split up in pre- and post-treatment periods and group by treated unit and control units yielding $\mathbf{Y}_{\text{t,pre}}^{\text{obs}}$ a T_0-vector, $\mathbf{Y}_{\text{c,pre}}^{\text{obs}}$ a $N \times T_0$ matrix (excluding the treated unit 0) and similarly, $\mathbf{Y}_{\text{t,post}}^{\text{obs}}$ and $\mathbf{Y}_{\text{c,post}}^{\text{obs}}$ for the post-treatment period, respectively. Combining these matrices yields:

$$\mathbf{Y}^{\text{obs}} = \begin{pmatrix} \mathbf{Y}_{\text{t,post}}^{\text{obs}} & \mathbf{Y}_{\text{c,post}}^{\text{obs}} \\ \mathbf{Y}_{\text{t,pre}}^{\text{obs}} & \mathbf{Y}_{\text{c,pre}}^{\text{obs}} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{\text{t,post}}(1) & \mathbf{Y}_{\text{c,post}}(0) \\ \mathbf{Y}_{\text{t,pre}}(0) & \mathbf{Y}_{\text{c,pre}}(0) \end{pmatrix} \tag{1.8}$$

The causal effect of interest in equation (1.5) depends on the pair of matrices $\mathbf{Y}_{\text{t,post}}(1)$ and $\mathbf{Y}_{\text{t,post}}(0)$. While $\mathbf{Y}_{\text{t,post}}(1)$ is observed, $\mathbf{Y}_{\text{t,post}}(0)$ represents the counterfactual and is therefore unobserved. The goal is now to impute the unobserved control outcomes for the treated unit $\mathbf{Y}_{\text{t,post}}(0)$ on the basis of the pre-treatment period outcomes for both treated and control units and the post-treatment outcomes for the control units, i.e. $\mathbf{Y}_{\text{t,pre}}(0)$, $\mathbf{Y}_{\text{c,pre}}(0)$ and $\mathbf{Y}_{\text{c,post}}(0)$. The question on the importance of covariates has been debated in the literature. Athey and Imbens (2017) and Doudchenko and Imbens (2017) argue that the outcomes tend to be substantially more important than covariates in term of predictive power. Therefore, to construct the synthetic control it is often sufficient to minimize the difference between the treated and control outcomes prior to the treatment. Furthermore, Kaul, Klößner, Pfeifer, and Schieler (2021) show that if all lagged outcomes are included in the synthetic control method as introduced by Abadie et al. (2010), covariates become redundant. In addition, in my particular analysis many covariates are already implicitly included in the outcomes. This is due to the fact that I use total returns and to form these, I need to take the coupon rate, coupon frequency and maturity into account. Liquidity, general industry and company stress as well as potential prior downgrades should also already be reflected in the price process. I therefore focus in my analysis on imputing the unobserved control outcomes for the treated unit solely on the basis of the three aforementioned outcomes $\mathbf{Y}_{\text{t,pre}}(0)$, $\mathbf{Y}_{\text{c,pre}}(0)$ and $\mathbf{Y}_{\text{c,post}}(0)$.

1.4.2 Identification

The estimation of the causal effect in equation (1.5) needs an identifying assumption since the counterfactual for the treated unit, $Y_{0,t}(0)$ for $t = T_0 + 1, \ldots, T$, is unobserved. The true counterfactual cannot be observed in general, therefore the validity of a particular identifying assumption cannot be tested empirically (Imbens & Wooldridge, 2009). In this paper, I use the conditional independence assumption for the identification of the treatment effect. The conditional independence assumption is discussed by Angrist and Pischke (2009). O'Neill, Kreif, Grieve, Sutton, and Sekhon (2016) and Kinn (2018) relate it to synthetic control methods. It is assumed that the potential outcome can be expressed by some function of an unobserved common time effect γ_t, unobserved time-varying unit specific effects $\phi_{i,t}$ and a vector of observed time-varying unit specific covariates $\mathbf{X}_{i,t}$, i.e. $Y_{i,t}(0) = f(\gamma_t, \phi_{i,t}, \mathbf{X}_{i,t})$. If the potential outcome $Y_{i,t}(0)$ was not exposed to a treatment, as in the case of the control units, it is independent of the treatment $W_{i,t}$ conditional on $\gamma_t, \phi_{i,t}$ and $\mathbf{X}_{i,t}$, i.e. $Y_{i,t}(0) \perp\!\!\!\perp W_{i,t} | (\gamma_t, \phi_{i,t}, \mathbf{X}_{i,t})$. The conditional independence on past outcomes assumption states that the potential outcome without treatment for both groups is the same in expectation conditional on past outcomes and observed covariates:

$$Y_{i,t}(0) \perp\!\!\!\perp W_{i,t} | (\mathbf{Y}^{\text{obs}}_{\text{i,pre}}, \mathbf{X}_{i,t}), \tag{1.9}$$

where $\mathbf{Y}^{\text{obs}}_{\text{i,pre}}$ is a vector of realized outcomes for unit i in the T_0 periods prior to the introduction of the treatment. The conditional independence assumption therefore proxies for the unobserved confounding factors γ_t and $\phi_{i,t}$ by conditioning on the full set of pre-treatment outcomes $\mathbf{Y}^{\text{obs}}_{\text{i,pre}}$. Under this assumption, it is anticipated that individuals with similar outcomes in the pre-treatment period would have similar potential treatment-free outcomes in post-treatment periods after conditioning on observed covariates $\mathbf{X}_{i,t}$. Since $Y_{i,t}(0)$ is affected by observed as well as unobserved confounders, Abadie et al. (2010) argue that units with similar outcomes in the pre-treatment period are also likely to have similar values of the time-varying unobserved confounding factors. Thus, if the synthetic control is able to find a weighted combination of controls that matches the treated unit, the time-varying unobserved confounding factors will also likely be balanced. In particular, in estimating the preannouncement sell-off fallen angel bonds suffer, I argue that before the sell-off starts, we can find a weighted combination of bonds that match the total return profile of the fallen angel bond. That weighted combination should in essence reflect similar bonds, bonds with similar values of the coupon rate, maturity, yield-to-maturity, level of the bond price, liquidity, industry stress, potential prior downgrades and rating profile. After the fallen angel bond has been pushed into a sell-off, we should be able to proxy the development of its total return in absence of the sell-off by using the weighted combination obtained from the periods prior to the sell-off.

Another assumption, that is implicitly stated in equation (1.7), is the stable unit treatment value assumption (SUTVA). It states that only two potential outcomes exist and that one of them is observed for each unit i. The SUTVA first appeared in Rubin (1980), but it had already been discussed in earlier studies. E.g. Cox (1958) assumes no interference between units, i.e. the potential outcome observation on one unit should be unaffected by the particular assignment of treatments to the other units. The SUTVA is therefore violated in the presence of general equilibrium effects (Heckman, Lalonde, & Smith, 1999), peer-effects, or in the presence of externalities and spillover effects. This is a very strict assumption and is unlikely to be satisfied on financial markets. A discussion of the causality of the estimated effects in my particular setting with the fallen angel event study is given in subsection 1.4.5.

1.4.3 Estimation

If we focus for ease of exposition only on time period T, the causal effect for unit 0 from equation (1.5) can be written as $\tau_{0,T} = Y_{0,T}(1) - Y_{0,T}(0)$. Considering that unit 0 receives the treatment in period T, $Y_{0,T}(1) = Y_{0,T}^{\text{obs}}$ with the causal effect therefore being $\tau_{0,T} = Y_{0,T}^{\text{obs}} - Y_{0,T}(0)$. $Y_{0,T}(0)$ is therefore the only unknown component and needs to be estimated. Doudchenko and Imbens (2017) note that many estimators commonly used in the literature share the following linear structure:

$$\hat{Y}_{0,T}(0) = \mu + \sum_{i=1}^{N} \omega_i Y_{i,T}^{\text{obs}}. \tag{1.10}$$

Therefore to impute the unknown $Y_{0,T}(0)$, a linear combination with intercept μ and weight ω_i for the outcome of control unit i in time T is proposed. One natural way to estimate this relationship might be by using ordinary least squares (OLS):

$$(\hat{\mu}^{\text{ols}}, \hat{\omega}^{\text{ols}}) = \arg\min_{\mu,\omega} \sum_{s=1}^{T_0} \left(Y_{0,s}^{\text{obs}} - \mu - \sum_{i=1}^{N} \omega_i Y_{i,s}^{\text{obs}} \right)^2. \tag{1.11}$$

The estimation of this equation involves T_0 observations and $N+1$ predictors (N controls and μ). However, in high-dimensional settings where $N+1 >> T_0$ some form of regularization of the weights ω is required to be able to estimate the regression. Doudchenko and Imbens (2017) propose the use of an elastic-net type penalty for the weights that combines the penalties from the least absolute shrinkage and selection operator (Lasso) and Ridge regression. Their objective function including the elastic-net penalty function is the following:

$$\begin{aligned}(\hat{\mu}^{\text{en}}(\lambda,\alpha), \hat{\omega}^{\text{en}}(\lambda,\alpha)) = \arg\min_{\mu,\omega} & \, ||\mathbf{Y}_{\text{t,pre}}^{\text{obs}} - \mu - \omega^T \mathbf{Y}_{\text{c,pre}}^{\text{obs}}||_2^2 \\ & + \lambda \left(\frac{1-\alpha}{2} ||\omega||_2^2 + \alpha ||\omega||_1 \right),\end{aligned} \tag{1.12}$$

where λ and α represent the overall shrinkage and the value of the elastic-net mixing parameter that determines the relative weight of Lasso vs. Ridge regularization, respectively. Doudchenko and Imbens (2017) do not normalize $\mathbf{Y}_{\text{c,pre}}^{\text{obs}}$, since this would change the weights when the variable is renormalized, thus affecting the construction of the synthetic control. Furthermore, the authors relax in their paper three assumptions common to traditional synthetic control methods as introduced by Abadie et al. (2010). These are

1) no-intercept, $\mu = 0$; 2) adding-up, $\sum_{i=1}^{N} \omega_i = 1$ and 3) non-negativity, $\omega_i \geq 0$, $i = 1, \ldots, N$. The DID method is relaxed with respect to the constant-weights assumption, $\omega_i = \bar{\omega}$, $i = 1, \ldots, N$ as well as the adding-up and the non-negativity assumption.

In order to determine the value of λ and α, I suggest to use $h\nu$-block cross-validation as proposed by Racine (2000). $h\nu$-block cross-validation is a consistent cross-validatory method to perform model-selection for dependent data. It consists of dividing the time-series of pre-treatment periods T_0 into K subsets, i.e. "folds," and because of the dependent nature of time-series one needs to further remove h observations from either side of the fold that serves as validation set. More formally, the training set has size $T_0 - 2h - \nu$, where $\nu = \lfloor T_0/K \rfloor$. For the purposes of this paper, I follow the rule-of-thumb suggestion of Burman, Chow, and Nolan (1994), setting $h = T_0/6$. According to Burman et al. (1994), $h = T_0/6$ appears to be a sensible choice in a variety of settings. Continuing with the cross-validation procedure, for a given λ and α one solves equation (1.12) by using the first training set and then computes the mean squared error on the validation set. This procedure is repeated K times, each time leaving out another validation set, such that all validation sets are left out once. The optimal pair $(\lambda_{\text{opt}}^{\text{en}}, \alpha_{\text{opt}}^{\text{en}})$ is chosen such that the average of the cross-validated errors is minimized. For practical purposes, often a finite set of values for $\alpha \in \{0, 0.05, 0.10, \ldots, 0.90, 0.95, 1\}$ and all possible values for $\lambda, \lambda \in (0, \infty)$ is considered. Once $(\lambda_{\text{opt}}^{\text{en}}, \alpha_{\text{opt}}^{\text{en}})$ have been determined, one can estimate equation (1.12) on the full T_0 to determine $(\hat{\mu}^{\text{en}}(\lambda_{\text{opt}}^{\text{en}}, \alpha_{\text{opt}}^{\text{en}}), \hat{\omega}^{\text{en}}(\lambda_{\text{opt}}^{\text{en}}, \alpha_{\text{opt}}^{\text{en}}))$. These estimates can then be used to construct the synthetic control and estimate the causal effect in equation (1.5).

In this application, I choose to only use the Lasso penalty function resulting in the following objective function:

$$\begin{aligned}(\hat{\mu}^{\text{lasso}}(\lambda), \hat{\omega}^{\text{lasso}}(\lambda)) = \arg \min_{\mu,\omega} &||\mathbf{Y}_{\text{t,pre}}^{\text{obs}} - \mu - \omega^T \mathbf{Y}_{\text{c,pre}}^{\text{obs}}||_2^2 \\ &+ \lambda \left(||\omega||_1\right), \end{aligned} \tag{1.13}$$

where $(\hat{\mu}^{\text{lasso}}(\lambda_{\text{opt}}^{\text{lasso}}), \hat{\omega}^{\text{lasso}}(\lambda_{\text{opt}}^{\text{lasso}}))$ now denote the estimates used to construct the synthetic control. The reason for only choosing the Lasso penalty function is that the elastic-net – due to the Ridge penalty component – does not often set weights exactly to zero. For prediction purposes this can make sense, but due to the causal question at hand, attaching non-zero weights to controls that are not similar to the fallen angel bond is rather counterintuitive.

1.4.4 Inference

Conducting inference for synthetic control methods is not an easy task. Main reasons are the lack of randomization and that no probabilistic sampling is used to choose sample units (Abadie et al., 2015). In their seminal study, Abadie et al. (2010) conduct inference on the estimated treatment effects by means of randomization inference. What randomization inference essentially does, is to assess if the estimated treatment effect of the treated unit is "abnormally" large in comparison to the treatment effects estimated from pseudo-treated units, i.e. placebo experiments. Pseudo-treated units are control units which are assigned the treated state – each control takes the role of a pseudo-treated once – and all other control units except the pseudo-treated unit are used as the control group. Conducting this placebo experiment yields a distribution of effects. Therefore, if one wants to claim some kind of "significance" of the treatment, the estimated treatment effect of the treated unit needs to lie outside or in the tails of the distribution.

Improving inference methods is an active field in the literature, with many contributions recently, see e.g. Ferman and Pinto (2019), Firpo and Possebom (2018), Xu (2017), Li (2020) and Chernozhukov, Wüthrich, and Zhu (2021). These new methods generally advocate some sort of bootstrapping approach where either the unit that is treated is viewed as random and/or that the period where the treated unit first receives the treatment is viewed as random.

For my specific application the question of interest would be inference on averages of estimated treatment effects and whether these are statistically significantly different from the cumulative abnormal returns as estimated from traditional event studies. Such kind of inference procedure has to my knowledge not yet been developed by the literature and would face many challenges due to the multiple layers in estimation uncertainty stemming from the synthetic control. Due to my analysis not being interested in estimating the size of the causal effect per se, but rather using it in order to make buying/selling decisions and analyzing subsequent average postannouncement returns, I do not conduct inference on the estimated causal effects. I do, however, conduct inference on average postannouncement returns, i.e. whether the difference in means between using the synthetic control vs. a different counterfactual is significant. I do this via the classical way of doing inference in traditional event studies, as described in section 1.2.

1.4.5 The question of causality in this particular application

After introducing a causal event study framework in section 1.4 and already discussing some potential pitfalls, one might ask, can the estimated preannouncement sell-off of fallen angel bonds be truly regarded as causal? The short answer is: Probably not. This is not in particular due to the introduced, general framework for a causal event study in section 1.4, but rather has to do with my particular research setting and question at hand. Specifically, whereas in comparative case studies there is a clear treatment at a clear date, e.g. implementation of a new policy or a change in a rule, I base my treatment on the outcome, i.e. the future downgrade event. More specifically, I assume that for bonds that are being downgraded, the preannouncement sell-off starts 100 days prior to the downgrade date. These 100 days are not simply plucked out of thin air or based on observing when in my particular sample fallen angels' bond prices start to drift off on average, but rather rely on prior literature such as Steiner and Heinke (2001), who already find that fallen angel bonds feature a negative excess return 100 days prior to the downgrade announcement. A later study conducted by Dor and Xu (2011), similarly finds up to three months prior to rerating announcements a negative price effect. Nevertheless, this is a very strong assumption and must not hold for the individual bond. Furthermore, the SUTVA is also likely to be violated due to the abundance of peer-effects, externalities and spillover effects on financial markets. Finally, the question is if it matters if the estimated preannouncement sell-off of fallen angel bonds is not causal in nature. The answer thereby depends on what the aim of the analysis is. If the aim of the paper would be to accurately estimate the size of the true causal impact of being downgraded to high-yield, it would absolutely be of importance, whether the assumptions can be regarded as satisfied to a high enough of a degree. If, however, as in this paper the aim of the analysis is rather simply using this estimated "causal" effect in order to make buying/selling decisions and investigating subsequent investment gains, the question of true causality becomes less of an issue.

1.5 Data

To analyze the preannouncement sell-off of fallen angel bonds I use a very comprehensive dataset from the OTC-market. The bond transaction tick data is available from July 2002 until December 2018[1] through Financial Industry Regulatory Authority's (FINRA) Trade Reporting and Compliance Engine (TRACE), which was introduced by the National Association of Securities Dealers (NASD) in 2002, in an effort to improve transparency in the OTC corporate bond market. From February 2005 onwards, 99 percent of all TRACE eligible bond transactions are covered on an intra-day basis. Information contained in the database is listed by transaction, key metrics include trade date, time, par volume, yield and transaction price. TRACE is therefore the most comprehensive source of pricing information when it comes to research questions concerning the U.S. corporate bond market. It is available on Wharton Research Data Services (WRDS). TRACE features two different versions, a standard and an enhanced version. The difference is that in the enhanced version all trades are included since 2002, whereas the standard edition only includes specific segments of bonds, i.e. until 2005, practically no high-yield trades were disclosed. Furthermore, the standard TRACE data includes a cap on the volume traded, whereas the enhanced edition reports the uncapped, exact volume. The disadvantage of the enhanced TRACE version is that it is reported with a lag, whereas the standard TRACE data is not. Fortunately, I am able to use the TRACE enhanced data for the full sample time. As discussed in section 1.2, using effective transaction prices and not some form of matrix prices is of utmost importance when analyzing relations between prices and ratings. The TRACE data, however, needs to be cleaned thoroughly before it can be used. Due to the fact that the TRACE data is dealer reported, errors can happen. Instead of correcting the data in the database directly, trade messages are appended, indicating either cancellations, corrections or reversals. Trades can also be double counted in the TRACE system, due to the fact that various parties can report the same trade. I follow the steps outlined in Dick-Nielsen (2009) and Dick-Nielsen (2014) to take care of the cancellation, correction, reversal and double counting issues. I followed Bessembinder et al. (2009) in calculating daily trade-weighted prices and followed Bai, Bali, and Wen (2019) in excluding trades with a par volume of less than 10,000 USD.

[1] I am constrained using data only up until December 2018, because my acquired Altman-Kuehne NYU Salomon Center Corporate Bond Default Master database ends then.

Table 1.1: **Sample setup**. In this table I describe the detailed steps I take to filter the TRACE dataset and the impact of those steps on the dimensions of my dataset.

	total sample			fallen angels sample	
	obs	issues	firms	events	firms
starting sample, TRACE raw data	22,063,238	173,017			
- remove trade records, price reversal > |10| USD	22,043,784	173,017			
- match bond CRSP link database	16,146,320	130,281	3,354		
- match Mergent FISD issue database	15,679,664	113,216	3,007		
- remove privately placed securities not available to the public		113,195			
- remove private placement issues (SEC Rule 144a)		112,410			
- remove bonds denominated in a foreign currency		112,399			
- remove bonds with country domicile outside U.S.		71,213			
- remove Yankee bonds		71,109			
- remove Canadian bonds		71,098			
- remove bonds not in "CZ," "UCID," "USBN," "CMTZ," "CMTN," "RNT," "CDEB" categories		68,266			
- remove asset backed bonds		68,261			
- remove convertible bonds		68,250			
- coupon type = "V" (variable coupon bonds)		52,661			
- coupon change indicator = "D," "R," "T" (coupon that is allowed to change)		52,495			
- remove perpetual bonds		52,493			
- remove preferred securities		48,523			
- missing values: private placement		48,446			
- missing values: security level		48,437			
- missing values: interest frequency		48,421			
- missing values: coupon		48,386			
- missing values: day count basis		48,362			
- missing values: coupon type		48,360			
- missing values: putable		48,356			
- conflict: interest frequency = 0 but coupon != 0		48,321			
- conflict: coupon = 0 but interest frequency != 0	11,918,157	48,319	1,981		
- match Mergent FISD ratings database	11,918,152	48,315	1,981	3,000	318
- remove 2nd fallen angel event (only keep the first fallen angel event for each bond)				2,940	318
- remove bonds due to insufficient maturity ($\leq$ 24 months from downgrade date)				1,853	280
- remove bonds due to maturity > 30 years				1,841	280
- remove bonds downgraded straight into default rating proximity (CCC+ or worse, see Table 1.2)				1,821	276
- match bond default database				1,821	276
- remove bonds that already defaulted prior downgrade date or in 10 days after				1,811	274
- remove bonds not traded during 1 year window around downgrade date				1,700	264
- remove bonds not traded during training window ($-365 : -101$, 265 days)				1,507	239
- remove bonds with insufficient training data (start data > -200, < 100 days max. available)				1,426	233
- remove bonds with insufficient number of training observations (< 10)				1,194	204
- remove bonds with insufficient number of testing ($-100 : -1$, 100 days) observations (< 10)				904	184
- remove bonds not traded in first 10 days post downgrade date				839	176

Table 1.1 displays the detailed sample setup. Firstly, I needed to delete trade records that featured a daily reversal in prices over 10 USD in absolute terms, because there were some obviously faulty trade prices. E.g. if a bond trades at 100 on Monday, drops to below 90 on Tuesday, and then features again a price of over 100 on Wednesday, I exclude the trade record for Tuesday for that particular bond. As evidenced by Table 1.1, only roughly 0.09% of trade records got removed and checking a large subset of them manually, confirmed that they were indeed outliers. I then need to match two databases to be able to uniquely determine the firm that issues the individual bonds (bond CRSP link database) and obtain information of the issue such as coupon, coupon frequency, day count convention, maturity (Mergent FISD issue database). I then follow the cleaning steps outlined by Bai et al. (2019), most importantly removing bonds in a different currency than the USD, bonds issued from firms outside the jurisdiction of the U.S., asset backed bonds, bonds with a variable coupon rate for which a total return computation is very cumbersome and prone to error,[2] perpetual bonds and preferred securities.[3] This big cleaning step reduced the number of trade records by 24% down to 11,918,157 resulting in having a sample of 48,319 issues of 1,981 firms.

The goal of the next step is to create a liquid sample of fallen angel bonds. I first match in rating information of the individual issues from the Mergent FISD ratings database. Table 1.2 displays an overview of the ratings the different rating agencies attach to individual bonds. The Mergent FISD ratings database includes the exact day of the rating changes from Fitch, S&P and Moody's. To define when a bond becomes a fallen angel, I use the index inclusion rules from Bloomberg Barclays benchmark indices. A bond is excluded from the index, if its middle rating drops below investment-grade. The middle rating is applied when all three rating agencies have an outstanding rating, if only two have an outstanding rating, the lower is the relevant rating for index inclusion. This way I can initially identify 3,000 fallen angel events of 318 firms. I then remove 60 fallen angel events, because I do not want multiple fallen angel events for individual bonds. These are mostly due to bonds that were about to expire and one of the rating agencies ceased their rating of the issue which led to jumps across the investment-grade, high-yield barrier. I then remove fallen angel events that after the downgrade announce-

[2] Representative bond benchmark indices such as Bloomberg Barclays also only include bonds with a fixed coupon rate.

[3] The bonds in this sample belong to one of the following categories: CZ-U.S. Corporate Zero, UCID-U.S. Corporate Insured Debenture, USBN-U.S. Corporate Bank Note, CMTZ-U.S. Corporate Medium Term Note (MTN) Zero, CMTN-U.S. Corporate MTN, RNT-Retail Note, CDEB-U.S. Corporate Debentures.

Table 1.2: **Ratings**. This table compiles and compares the ratings given by the three rating agencies Fitch, S&P and Moody's. A previously investment-grade rated bond crossing the investment-grade/high-yield barrier is considered a fallen angel bond.

Fitch	S&P	Moody's		Rating grade description (Moody's)
AAA	AAA	Aaa	IG	Minimal credit risk
AA+	AA+	Aa1		Very low credit risk
AA	AA	Aa2		
AA-	AA-	Aa3		
A+	A+	A1		Low credit risk
A	A	A2		
A-	A-	A3		
BBB+	BBB+	Baa1		Moderate credit risk
BBB	BBB	Baa2		
BBB-	BBB-	Baa3		
BB+	BB+	Ba1	HY	Substantial credit risk
BB	BB	Ba2		
BB-	BB-	Ba3		
B+	B+	B1		High credit risk
B	B	B2		
B-	B-	B3		
CCC+	CCC+	Caa1		Very high credit risk
CCC	CCC	Caa2		
CCC-	CCC-	Caa3		
CC	CC	Ca		In or near default, with possibility of recovery
C	C			
DDD	SD	C		In default, with little chance of recovery
DD	D			
D				

ment have less or equal to 24 months left until redemption. This is due to the fact, that index exclusion rules for major benchmark indices require a minimum time-to-maturity of 1 year. Keeping a bond with a short maturity, could suffer a mechanical price pressure due to being excluded from the index. I also proxy for perpetual bonds by excluding fallen angel bonds with a remaining maturity of over 30 years. I exclude bonds rated directly to CCC+ or worse, because these include issues that went straight into default at the time of the downgrade, e.g. Lehman Brothers, and an investor would never consider buying them. In order to control for defaulting bonds directly, I acquired the Altman-Kuehne NYU Salomon Center Corporate Bond Default

Master database. Defaults are kept in the dataset to avoid any survivorship bias. Their price, however, changes only according to their trading price (which represents the recovery rate). Defaulted issues are assumed not to pay coupons anymore and therefore their traded price equals their total return. In my analysis, I only start to invest in bonds after 10 days postannouncement. This is due to the fact that immediately after the downgrade, fallen angels tend to suffer a substantial price shock as outlined in section 1.2. I therefore remove bonds that already featured a default event[4] up to 10 days postannouncement.

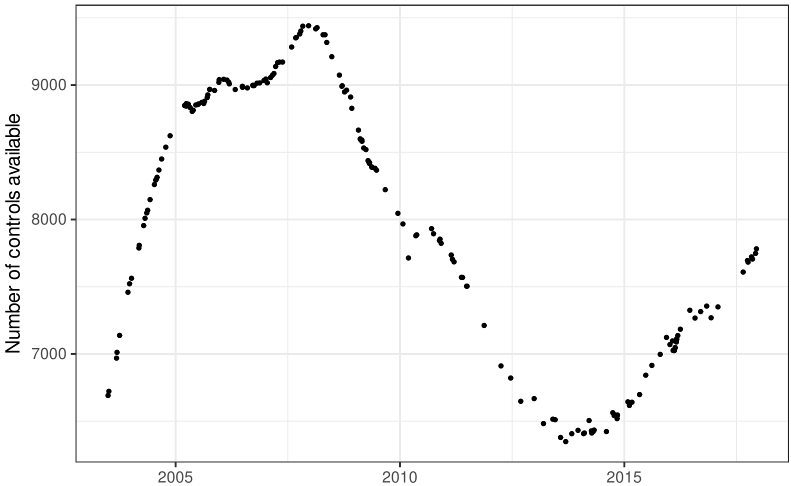
The next couple of cleaning steps deal with having enough trades in order to estimate the synthetic control as well as preserving liquidity of the fallen angel bonds around the downgrade date. In total, I cover data 1 year around the downgrade event. The downgrade event is day 0. The event window therefore goes from day -365 to day +365, covering 731 days in total. The training window where I estimate the Lasso, covers 265 days in total, from days -365 to -101. The testing window, where I use the weights estimated using the Lasso on the training window and then fit the model on the testing data in order to get the synthetic control counterfactual, covers 100 days in total, from days -100 to -1. This testing window period I coin in this paper the "preannouncement." A couple of fallen angel bonds were not traded 1 year around their downgrade date as well as during the training window and thus needed to be removed from the sample. Fallen angel bonds not traded at least once prior to day -200, or having less than 10 trade records in the training/testing window were removed as well. Finally, I also removed bonds not traded in the first 10 days post their downgrade date.

The final sample includes 839 fallen angel bonds of 176 firms. The fallen angel sample featured 70 defaulted bonds of 23 firms. In general, if a bond is not traded on a given day, its prior price is filled forwards to the next available transaction price. The bond, however, accumulates accrued interest. Due to the Mergent FISD issue database, I am able to calculate the exact total return using equation (1.2), because I have full data on the coupon, coupon frequency, day count convention, last interest date and maturity.

Figure 1.1 displays for each of the 182 distinct fallen angel downgrade event dates the number of potential control bonds, i.e. investment-grade rated

[4] The Altman-Kuehne NYU Salomon Center Corporate Bond Default Master database defines a default as being either a bankruptcy, missed interest payment or distressed exchange.

Figure 1.1: **Number of IG rated issues**. This figure shows how many control bonds were available for each of the 182 distinct fallen angel downgrade events. Control bonds needed to maintain an investment-grade rating from day -365 to day +10 and had to have a cumulative total return from day -365 to day -1 of less than 100%.



bonds that can be used to train and fit the synthetic control 1 year up to the event. The number of investment-grade issues outstanding varied substantially over the sample period, reaching a peak shortly before the onset of the global financial crisis in 2008 of around 9,500 issues and then steadily decreased to under 8,000 issues. In order for the individual control bond to qualify to be included in the sample, it had to exhibit a continuous investment-grade rating from days -365 through to day +10. Furthermore, outliers were removed, by removing control bonds that featured a doubling or more of their indexed total return value from 100 (day -365) to 200 (day -1) or over.

Table 1.3 shows in which years and by how many firms the 839 fallen angel events occurred. There is a clustering of fallen angel events in the years from 2004 until 2009 and from 2014 until 2016. In particular in 2005, there were 285 fallen angel events. From those 285 fallen angel events, however, 236 can be directly attributed to "Ford Motor Company" being downgraded to high-yield. The period between 2010 and 2013 features comparatively few fallen angel events. The example of having 236 fallen angel bonds of the same firm, highlights the importance of grouping individual fallen angel bonds by firm

Table 1.3: **Number of fallen angel events**. This table displays in which year, how many fallen angel events occurred by how many firms. There is a large clustering of 236 events in 2005 from "Ford Motor Company."

Year	Bonds	Firms
2003	14	7
2004	48	17
2005	285	19
2006	47	19
2007	45	18
2008	84	12
2009	97	23
2010	24	7
2011	13	8
2012	7	5
2013	26	6
2014	66	13
2015	21	8
2016	48	19
2017	14	9
	839	

and taking averages.

Figure 1.2 displays indexed fallen angel bonds' total return (grouped by firm with an equal-weighting scheme). Figure 1.3 features the notorious V-curve, that is typical for the price behavior of fallen angel bonds. Figure 1.3 is generated by first grouping fallen angels bonds' total return by firm (average returns by firm) and then grouping by day (average returns by day) and then indexing it to a base of 100. From Figure 1.3 it is evident, that around 2 percent of the value would temporarily be lost a few days after the downgrade event. That the minimum on average is slightly after the downgrade announcement, can be explained by the reaction time of investors and passive investment vehicles to the downgrade announcement. For instance, some investors might not have anticipated the downgrade and after hearing about the news of the rerating by the rating agency, need to sell the fallen angel bond due to their investment guidelines. Furthermore, a passive investment vehicle such as an exchange traded fund typically rebalances its portfolio holdings at the end of the month. Figure 1.3 also shows, that 100 days prior to the downgrade an-

Figure 1.2: **Fallen angel return (grouped by firm)**. This figure shows indexed fallen angel bonds' total return. It is generated by grouping fallen angel bonds' total return by firm (average return by firm) and then indexing it to a base of 100.

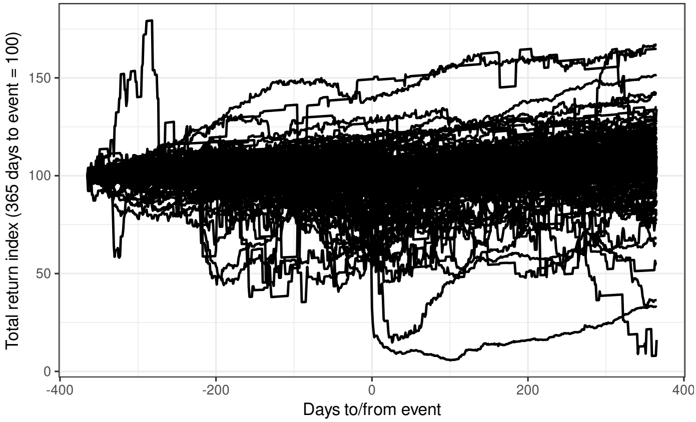
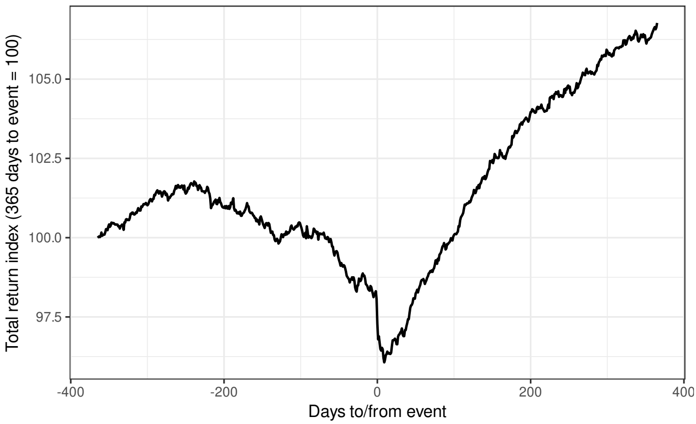


Figure 1.3: **V-curve, mean fallen angel return (grouped by firm)**. This figure shows the notorious V-curve, that is typical for the price behavior of fallen angel bonds. It is generated by first grouping fallen angels' bond total return by firm (average returns by firm) and then grouping by day (average returns by day) and then indexing it to a base of 100.



nouncement, the kink in the performance apparently starts. It can therefore be presumed, that at roughly 100 days before the downgrade announcement, investors engage in the sell-off. Next to the findings already presented by Steiner and Heinke (2001) and Dor and Xu (2011), this further motivates my choice for the 100 days prior to the downgrade announcement to be my preannouncement sell-off period. Data up until 100 days prior to the fallen angel event are therefore used to train the synthetic control. The synthetic control is then fit on the 100 days prior to the downgrade announcement to obtain the counterfactual. In terms of the time periods the causal event study framework uses in section 1.4, $1, \ldots, T_0$ corresponds to days -365 until -101 prior to the event, and $T_0 + 1, \ldots, T$ corresponds to days -100 until -1 prior to the event. T_0 would therefore mean that the pre-treatment period (before the preannouncement sell-off starts) amounts to 265 days in total. T on the other hand, would amount to 365 days, just before the downgrade to high-yield announcement by the rating agency on day 0. Revisiting the number of controls available in Figure 1.1, we can observe that for each distinct rating date $N+1 >> T_0$. We are therefore in the high-dimensional setting and some form of regularization of the weights ω is necessary, as discussed in subsection 1.4.3.

The task in the next section will be to estimate the abnormal preannouncement return. This will be done by training the synthetic control on the pre-treatment sample $1, \ldots, T_0$ and then fitting it on the treatment sample $T_0 + 1, \ldots, T$. The total return of fallen angel bonds will then subtract their individual synthetic controls' returns to form abnormal returns (equation (1.1)). These abnormal returns are then log-cumulated (equation (1.3)) and then grouped by firm with an equal-weighting scheme (equation (1.4)). The resulting vector $CAR_{(-100,-1),i}$ is therefore simply the treatment effect for a fallen angel bond – i.e. unit 0 in section 1.4 – in time period T, $\tau_{0,T} = Y_{0,T}(1) - Y_{0,T}(0)$ or equivalently due to $Y_{0,T}(1)$ being observed, $\tau_{0,T} = Y_{0,T}^{\text{obs}} - Y_{0,T}(0)$. Both the traditional event study setup in section 1.3 and the introduced causal event study setup in section 1.4 can therefore easily be related to each other. Investment applications using the size of the abnormal preannouncement return with the synthetic control as its counterfactual are considered and its benefits vs. using a "naive" broad-based investment-grade rated benchmark are highlighted.

1.6 Results

This section aims to estimate the preannouncement sell-off fallen angel bonds suffer and to use this estimate in order to be able to make better informed buying/selling decisions post-downgrade. The question hereby is, whether a custom benchmark – the constructed synthetic control – can provide investors with some added utility, rather than simply using a broad-based investment-grade bond index as a counterfactual to gauge the fallen angel bond's potential cumulative abnormal preannouncement return. To achieve this aim, a synthetic control needs to be constructed in the pre-treatment period, i.e. before the sell-off of worried fund managers starts. This synthetic control should be similar to the fallen angel bond in all important aspects. This will allow the imputation of the unknown counterfactual for the fallen angel bond and thus to gauge the preannouncement sell-off on the fallen angel's bond price. The estimate of this preannouncement sell-off is an important metric to gauge the degree of overselling, which depresses the price of a fallen angel bond below its fair value. Why should this fair value of a bond be important? The fair value should be able to help investors' judgment about whether the price decline reflects overall market risk for these type of bonds or is purely driven as a result of a sell-off of a particular company and thus trying to avoid having a high-yield rated bond in the portfolio. If one further believes that a bond's fair value – i.e. counterfactual – can badly be proxied by an overall broad investment-grade market index that is the same for all bonds, one should consider trying to estimate a customized counterfactual, that arguably should fit the characteristics of the soon-to-be fallen angel bond better and thus should reflect a more credible counterfactual.

This has important implications on constructing fallen angel factor portfolios targeted on harvesting this anomaly resulting from institutional and behavioral factors. On the one hand, one could e.g. advocate a strategy that buys (i.e. goes long) bonds that trade at a discount to their fair value and sells (i.e. goes short) bonds that trade at a premium. Due to the fact that corporate bonds are very costly to short, a long only strategy that only invests in bonds that trade below their estimated fair value will be the focus of this paper. On the other hand, guidance could also be given to portfolio managers with respect to a possible weighting scheme for their potential fallen angels investment, that is, that it potentially could be profitable to overweight (more recently downgraded) fallen angel bonds that exhibit a larger discount to their fair value. Questions of how long a fallen angel bond should be included in the portfolio can be addressed by examining for various holding periods the excess returns with respect to a high-yield broad benchmark index. E.g., if fallen angel bonds would on average underperform a high-yield benchmark index for a holding period of 6 months, it would not make sense to take on the added risk of holding these fallen angel bonds and the investor should simply consider investing directly into a diversified high-yield market vehicle.

As already mentioned in section 1.5, the pre-treatment period, i.e. before the preannouncement sell-off started, is set to $T_0 = 265$ days. Since $N + 1 >> T_0$ is true for each fallen angel sample (see Figure 1.1), some kind of regularization of the weights ω needs to be applied. In this paper, I follow largely the approach of Doudchenko and Imbens (2017) with the exception of the choice of the penalty function. In particular, they propose a generalized synthetic control method which can cope with the high-dimensional case by applying an elastic-net type penalty on the weights ω. In my particular setting, however, a Lasso only type penalty on the weights makes more sense. The approach has been discussed in subsection 1.4.3. I choose to use $h\nu$-block cross-validation as proposed by Racine (2000) in order to tune the tuning parameter λ. What still needs to be set are the number of folds K for the cross-validation procedure. By doing this, one also needs to consider the rule-of-thumb suggestion from Burman et al. (1994) for setting the h in the $h\nu$-block cross-validation procedure. To recall, the h suggested by Burman et al. (1994) amounts to $h = T_0/6$. This means that $265/6 \approx 44$ observations get removed from either side of the fold that serves as validation set. If e.g., we choose to set $K = 3$, the first validation set includes the first 88 observations of T_0, the next 44 are omitted from the training set because they "belong" to the h. The training set therefore has a size of $T_0 - T_0/6 - \lfloor T_0/K \rfloor = 265 - 265/6 - 88 = 133$. Similarly, the third training set also amounts to 133 observations. The second, however, amounts to $T_0 - 2 \times T_0/6 - \lfloor T_0/K \rfloor = 265 - 2 \times 265/6 - 88 = 89$, because we have to remove h twice from either side of the validation set because it is in the middle. Note that for $K > 5$, entire folds get skipped over on either side of the validation set because of the h observations that need to be removed in the $h\nu$-block cross-validation procedure. I therefore propose to use $K = 3$ in this analysis.

Next, I estimate equation (1.13) with the above-mentioned cross-validation procedure to tune the tuning parameter λ. Note that I follow Doudchenko and Imbens (2017) with not imposing the traditional synthetic control method restrictions on the ω, as well as not normalizing the $\mathbf{Y}^{\text{obs}}_{\text{c,pre}}$. Figure 1.4 displays the percentage of control bonds selected for each fallen angel bond, based on the optimal cross-validated value of the $\lambda^{\text{lasso}}_{\text{opt}}$. On average, roughly 24 controls were selected for each fallen angel bond (0.3 percent of the sample size). For each fallen angel bond, there were on average 8,339 controls to choose from. With my estimates of $(\hat{\mu}^{\text{lasso}}(\lambda^{\text{lasso}}_{\text{opt}}), \hat{\omega}^{\text{lasso}}(\lambda^{\text{lasso}}_{\text{opt}}))$ I can build the synthetic control for each fallen angel bond. Figure 1.5 shows an example of two fallen angel bonds' total return index along with their respective estimated

Figure 1.4: **Number of controls chosen by Lasso**. This figure displays the percentage of control bonds selected for each fallen angel bond, based on the optimal cross-validated value of the $\lambda^{\text{lasso}}_{\text{opt}}$. 24 controls (0.3 percent of the sample size) were selected on average.

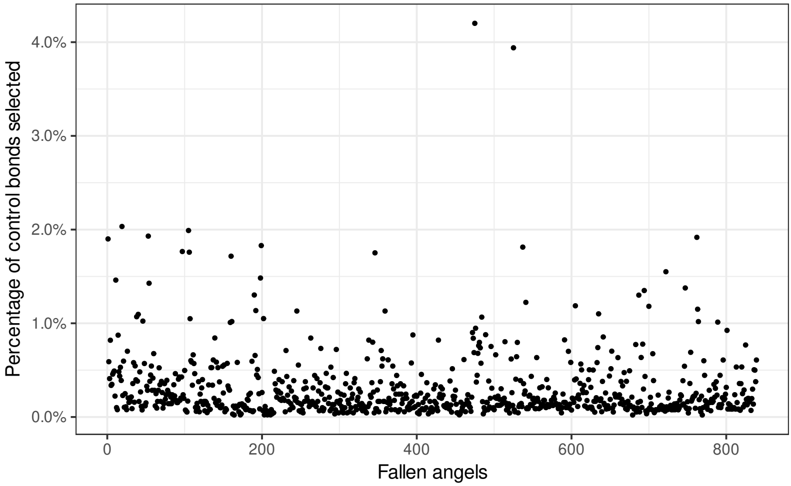
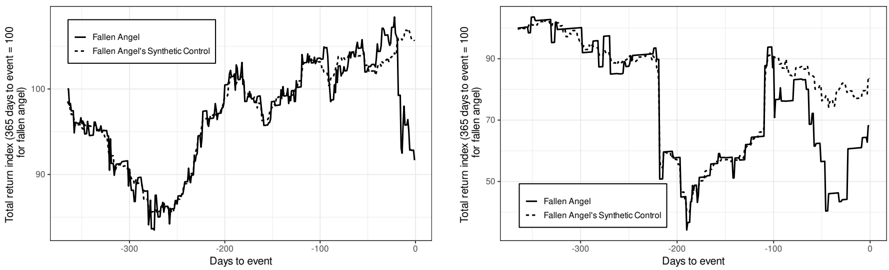


Figure 1.5: **Fallen angel bond and its estimated synthetic control**. This figure displays two fallen angels' total return index along with their estimated synthetic control. As can be seen from Figure 1.5b, one advantage of using the synthetic control is that it is also reasonably able to cope with less liquid issues.



(a) Cusip 12527GAE3

(b) Cusip 12557WNS6

synthetic control. One can see, that one further advantage of the synthetic control is, that it is also reasonably able to cope with less liquid issues, such as in Figure 1.5b. Following the estimation of the synthetic control, I take returns and plug them in equation (1.5) to obtain the "causal effect" $\tau_{0,t}$ for time periods $t = T_0 + 1, \ldots, T$. Note that that this equation corresponds to equation (1.1) for abnormal returns. Specifically, I subtract from the total returns of the fallen angel bond the returns of the estimated synthetic control. These abnormal returns are then log-cumulated (equation (1.3)) and then – following Grier and Katz (1976) and Bolognesi et al. (2014) – grouped by firm with an equal-weighting scheme (equation (1.4)). This yields the cumulated, aggregated, abnormal preannouncement return for time period T, i.e. the total size of the preannouncement sell-off of fallen angel bonds from a particular firm.

Figure 1.6: **Abnormal fallen angel preannouncement return (grouped by firm)**. This figure displays a histogram of the abnormal preannouncement return (using the synthetic control as counterfactual), i.e. the preannouncemet sell-off fallen angel bonds suffer. The abnormal preannouncement return is generated by cumulating the log-difference of fallen angel total returns and returns of the synthetic control, grouping by firm and then taking the average.

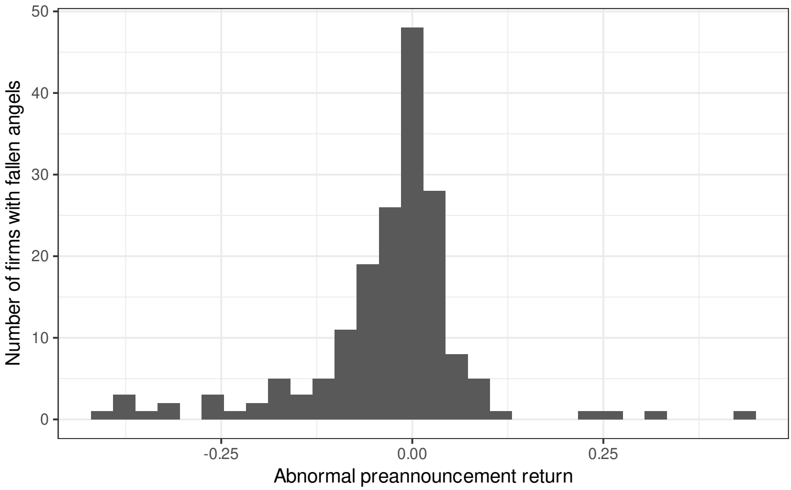


Figure 1.6 displays a histogram of the estimated abnormal preannouncement returns. If the abnormal preannouncement return is negative (positive), this means that the fallen angel bond trades at a discount (premium) relative to its estimated synthetic control. 63 percent of fallen angel bonds trade at a discount, 1 day before the rating agency announces their downgrade to high-yield and therefore removing them from the investment-grade universe. The average preannouncement sell-off of fallen angel bonds amounts to -3.46 percent.

However, when instead of the synthetic control all investment-grade issues (used in the construction of the synthetic control, i.e. the controls) are equal-weighted and used as the counterfactual, the preannouncement sell-off is more pronounced with a value of -3.75 percent. Value-weighting the controls yields a similar sell-off of -3.60 percent. The obtained results are as hypothesized, i.e. the sell-off is less pronounced when using a synthetic control as a counterfactual. It makes intuitively sense, that the preannouncement sell-off of fallen angels using the estimated synthetic control is less pronounced compared to simply using broad investment-grade bond indices. It can be argued that this is due to the synthetic control providing a counterfactual that is more similar to the fallen angel bond. Bonds that are about to be downgraded might experience a lower return than that of a broad investment-grade benchmark. This lower return results in a difference that is less negative when forming abnormal returns, thus yielding an overall less negative estimate of the preannouncement sell-off. Using a traditional event study therefore could tend to overestimate the size of preannouncement sell-off of fallen angel bonds, whereas using a framework inspired from causal inference could result in a more sensible estimate of the preannouncement sell-off by using a custom, more similar benchmark to the fallen angel bonds. In Appendix 1.A, Figures 1.8 and 1.9 show averaged returns for fallen angels and counterfactuals up until the downgrade announcement. In particular, it can be seen that the synthetic control was able to match the total return of fallen angels on average in the pre-treatment period. Figure 1.8 shows, that the equal-weighted investment-grade controls performed the best, followed by their value-weighted version and lastly the Bloomberg value-weighted benchmark index. This makes intuitively sense, because equal-weighted indices in general tend to fare better than indices with more concentrated weights (smaller, more risky and volatile, perhaps profitable issues are also included in equal-weighting). If as in Figure 1.9 counterfactuals are indexed at the start of the treatment-period, i.e. then when the preannouncement sell-off is about to happen, one can observe that the Bloomberg benchmark index actually features an overall lower return than the synthetic control. This leads to an average preannouncement sell-off of -3.40 percent that is slightly less negative than that of the synthetic control and would therefore counter our above train of thought about the hypothesized sizes and orderings of sell-offs under different counterfactuals. I attribute this, however, to the different rebalancing (monthly vs. daily) of the Bloomberg index and that it has not been calculated based on the TRACE dataset from scratch, thus being perhaps tilted towards specific big companies and not the entire universe of bonds. Nevertheless, the Bloomberg index is of utmost importance of being included as a counterfactual, because this is what investors have at hand when evaluating investment decisions and what has been used in the past in traditional event studies.

What is the use of estimating the preannouncement sell-off of fallen angels? By examining the V-curve in Figure 1.3, one might possibly argue, that fallen angel bonds with a larger preannouncement sell-off might subsequently generate higher returns. The question hereby is, can the synthetic control as a counterfactual better distinguish between over- and underpriced fallen angel bonds than broad investment-grade market indices? If it can better distinguish between over- and underpriced fallen angel bonds, more pronounced discounted fallen angel bonds should subsequently generate higher returns. Fallen angel bonds trading at a rather small discount or even above their fair value should therefore generate subsequently lower returns. There should therefore exist a negative relationship between the size of the preannouncement sell-off of fallen angel bonds and their subsequent postannouncement return. This can be formalized in the following regression:

$$CR_{(+11,+\text{Holding period}),j} = \alpha + \beta CAR_{(-100,-1),j} + \epsilon, \tag{1.14}$$

where $CR_{(+11,+\text{Holding period}),j}$ denotes the cumulative return of firm j's outstanding fallen angel bonds if bought 11 days after the downgrade announcement and held for a specified holding period. The constant is denoted by α and ϵ is the error term. The preannouncement sell-off of fallen angel bonds from firm j is given by $CAR_{(-100,-1),j}$. β denotes the coefficient on the preannouncement sell-off of a fallen angel bond, i.e. how strong the relation is between the size of the preannouncement sell-off and its subsequent postannouncement return over the holding period. If the synthetic control can better identify between over- and underpriced fallen angel bonds, its β should be more negative than that of the counterfactual that simply uses a broad investment-grade index.

In Table 1.4, the results of the estimation of equation (1.14) are shown. Results are tabulated for 4 different counterfactuals: 1) synthetic control, 2) all investment-grade bonds equally-weighted (calculated from the control sample used to construct the synthetic control), 3) all investment-grade bonds value-weighted and 4) as a robustness check the commonly used broad investment-grade bond index from Bloomberg. The results show, that when using the synthetic control as the counterfactual for fallen angel bonds, the estimated β is more negative for all holding periods, no matter whether equal-, value-weighted or even the Bloomberg index is used as counterfactual. This indicates, that using the synthetic control as a counterfactual

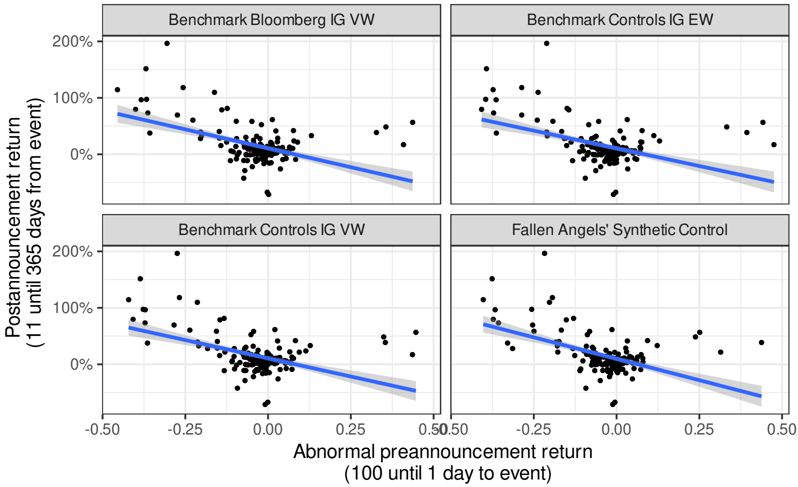
Table 1.4: **Regression results**. This table shows the results from regressions of postannouncement returns on the preannouncement sell-off of fallen angel bonds for 4 different counterfactuals and various holding periods. *, **, ***, denotes significance on the 10, 5, 1 percent level, respectively.

\Counterfactual	Synthetic Control		Controls IG EW		Controls IG VW		Bloomberg IG VW	
Holding period	β	Std. error	β	Std. error	β	Std. error	β	Std. error
1 month	-0.05	0.06	-0.03	0.05	-0.01	0.05	-0.01	0.06
2 months	-0.27	0.07***	-0.25	0.06***	-0.23	0.06***	-0.22	0.06***
3 months	-0.32	0.08***	-0.27	0.07***	-0.28	0.07***	-0.29	0.07***
4 months	-0.49	0.09***	-0.38	0.09***	-0.40	0.09***	-0.42	0.09***
5 months	-0.60	0.11***	-0.46	0.10***	-0.49	0.10***	-0.50	0.10***
6 months	-0.83	0.12***	-0.69	0.11***	-0.71	0.11***	-0.74	0.11***
7 months	-0.94	0.13***	-0.79	0.12***	-0.82	0.12***	-0.85	0.12***
8 months	-0.94	0.14***	-0.77	0.13***	-0.81	0.13***	-0.86	0.13***
9 months	-1.10	0.16***	-0.93	0.14***	-0.96	0.14***	-1.01	0.14***
10 months	-1.26	0.17***	-0.98	0.16***	-1.03	0.16***	-1.08	0.16***
11 months	-1.38	0.18***	-1.11	0.17***	-1.15	0.17***	-1.20	0.17***
12 months	-1.52	0.20***	-1.25	0.18***	-1.29	0.18***	-1.34	0.18***

can better distinguish between over- and underpriced fallen angel bonds than broad investment-grade market indices. Fallen angel bonds with a higher preannouncement sell-off tend to subsequently generate higher returns. Fallen angel bonds trading at a rather small discount or even above their fair value tend to generate subsequently lower (or even negative) returns.

Figure 1.7 displays the relationship between a firm's abnormal preannouncement return, calculated using the 4 different counterfactuals, and its subsequent 12-month postannouncement return. The β is simply the slope of the regression line in equation 1.14. Although the 4 scatterplots look fairly similar, it can be seen that when using the synthetic control as the counterfactual, the regression line is steeper. In Table 1.4 the estimated β coefficients are more negative, the longer the holding period. This indicates that the return of holding fallen angel bonds gets more positive, the longer the holding period is. This is in line with the general V-curve as displayed in Figure 1.3. Furthermore, Figure 1.7 shows that following a strategy that involves shorting bonds will not lead to superior investment results. This is directly visible as evidenced by the lack of observations in the lower-right quadrant, i.e. fallen angel bonds that traded at a premium with respect to their counterfactual and that generated subsequent negative returns. Fallen angels that traded at a premium therefore did not see a reversal in prices strong enough that would yield a profitable investment strategy. Employing a long-only investment ap-

Figure 1.7: **Abnormal preannouncement vs. 12-month postannouncement return (grouped by firm)**. This figure shows the relationship between a firm's average abnormal preannouncement and average postannouncement return if its fallen angel bonds are held for 12 months. Abnormal preannouncement returns are calculated using 4 different counterfactuals.



proach therefore seems to be the natural choice.

A natural starting point when wanting to investigate the economic benefits of investing in fallen angel bonds is: 1) If fallen angel bonds are bought, how much money can one make after n months on average and 2) is it worth taking on the additional risk of investing in fallen angel bonds or should one simply stick with holding risky high-yield, junk-rated bonds? Table 1.5 displays average equal-weighted returns by firm (176 firms in total) of buying fallen angel bonds – irrespective the size of their preannouncement sell-off – 11 days after they were downgraded and held for n months. Furthermore, the average outperformance vs. the Bloomberg high-yield bond benchmark index is also investigated. This naive fallen angel buying strategy yields a return of 14.81 percent after 1 year on average, with an outperformance of 4.26 percent. The outperformance, however, peaks after 10 months, suggesting that after that holding period, it is not profitable to hold fallen angel bonds anymore and that investors should rather switch to a diversified high-yield benchmark index.

Table 1.5: **Naive fallen angel buying strategy**. This table displays the average total return and outperformance vs. the Bloomberg high-yield bond benchmark index for various holding periods if fallen angels are bought no matter their preannouncement sell-off.

Holding period	return (%)	return - HY (%)
1 month	0.80	0.06
2 months	3.03	1.11
3 months	4.23	1.04
4 months	6.48	1.97
5 months	8.32	3.07
6 months	9.68	3.44
7 months	10.69	3.55
8 months	11.35	3.31
9 months	12.38	3.74
10 months	13.38	4.41
11 months	14.09	4.17
12 months	14.81	4.26

The analysis in Table 1.4 showed, however, that bonds that suffered a higher preannouncement sell-off, on average experienced a larger postannouncement return. This would suggest, that one should only buy fallen angels that trade at a discount with respect to their counterfactual and that the bigger this sell-off was, the bigger potential profit opportunities are.

Panel A in Table 1.6 shows the average total return and outperformance vs. the Bloomberg high-yield bond benchmark index for holding periods ranging from 1 through 12 months if fallen angels are bought if they trade at a discount vs. their counterfactual (abnormal preannouncement return $<$ 0). Results are tabulated for each considered counterfactual. Furthermore, Panel B in Table 1.6 restricts the choice of buying fallen angels only to the ones that featured the most severe preannouncement sell-off (abnormal preannouncement return $<$ 20th percentile of abnormal preannouncement return for each respective counterfactual). As in previous tables, individual fallen angels' bond return was averaged by firm using an equal-weighting scheme. Using the synthetic control as a counterfactual and only buying bonds that trade at a discount, results in an average total return of 18.71 percent if fallen angels are bought 11 days after the downgrade and held for 12 months. This total return is higher than the one which does not take into account the size of the preannouncement sell-off (see Table 1.5, 14.81 percent). Using the size of the preannouncement sell-off can therefore enhance the total return of

Table 1.6: **Fallen angel buying strategy**. This table displays the average total return and outperformance vs. the Bloomberg high-yield bond benchmark index for various holding periods if fallen angels are bought according to if their abnormal preannouncement return < 0 (Panel A), or if their abnormal preannouncement return $<$ 20th percentile of abnormal preannouncement return (Panel B). Abnormal preannouncement returns are calculated using 4 different counterfactuals, 1) synthetic control, 2) controls equally-weighted, 3) controls value-weighted and 4) using the industry standard Bloomberg investment-grade value-weighted benchmark index.

\Counterfactual	Synthetic Control		Controls IG EW		Controls IG VW		Bloomberg IG VW	
Holding period	return (%)	return - HY (%)	return (%)	return - HY (%)	return (%)	return - HY (%)	return (%)	return - HY (%)
Panel A: Long (abnormal preannouncement return < 0)								
1 month	0.78	0.04	0.53	-0.21	0.50	-0.24	0.59	-0.15
2 months	3.91	1.99	3.57	1.65	3.56	1.64	3.44	1.52
3 months	5.16	1.97	4.82	1.63	4.83	1.64	4.77	1.58
4 months	7.83	3.32	7.36	2.85	7.34	2.83	7.08	2.57
5 months	9.85	4.60	9.18	3.93	9.34	4.09	9.14	3.89
6 months	12.29	6.05	11.12	4.88	11.48	5.24	11.32	5.08
7 months	13.73	6.59	12.17	5.03	12.53	5.39	12.52	5.38
8 months	14.16	6.12	12.63	4.59	12.98	4.94	12.97	4.93
9 months	15.79	7.15	13.91	5.27	14.21	5.57	14.22	5.58
10 months	17.03	8.06	14.67	5.70	14.99	6.02	15.66	6.69
11 months	17.82	7.90	15.48	5.56	15.76	5.84	16.96	7.04
12 months	18.71	8.16	16.46	5.91	16.74	6.19	17.98	7.43
Panel B: Long (abnormal preannouncement return $<$ 20th percentile of abnormal preannouncement return)								
1 month	3.34	2.60	2.62	1.88	2.43	1.69	2.09	1.35
2 months	10.18	8.26	9.50	7.58	9.14	7.22	8.60	6.68
3 months	13.15	9.96	12.71	9.52	12.01	8.82	11.48	8.29
4 months	19.79	15.28	19.26	14.75	18.43	13.92	17.89	13.38
5 months	25.17	19.92	24.45	19.20	23.62	18.37	22.91	17.66
6 months	30.65	24.41	29.38	23.14	28.63	22.39	28.12	21.88
7 months	34.61	27.47	32.47	25.33	31.63	24.49	31.20	24.06
8 months	35.18	27.14	33.62	25.58	32.66	24.62	32.05	24.01
9 months	39.81	31.17	37.34	28.70	36.51	27.87	35.60	26.96
10 months	45.05	36.08	41.97	33.00	40.89	31.92	39.86	30.89
11 months	48.11	38.19	45.58	35.66	44.26	34.34	43.39	33.47
12 months	51.83	41.28	49.42	38.87	47.46	36.91	46.06	35.51

the investment by roughly 4 percentage points. Furthermore, the results in Panel A in Table 1.6 show, that using the synthetic control as counterfactual outperforms every other choice of counterfactual for any of the considered holding periods. For the 12-month holding period differences in total return using the synthetic control vs. a different choice of counterfactual range from -0.73 percentage points (Bloomberg investment-grade index) to -2.25 percentage points (equal-weighted investment-grade bonds). There is therefore an advantage of using the synthetic control as opposed to a simple benchmark. This advantage is even more pronounced, if only the fallen angels with the biggest preannouncement sell-off are considered (Panel B in Table 1.6). Differences in 12-month total return now range from -2.41 percentage points (equal-weighted investment-grade bonds) up until a whopping -5.77 percentage points (Bloomberg investment-grade index). For investors this has severe implications, because most of them only have access to the readily available Bloomberg investment-grade index for evaluating the size of the preannouncement sell-off.

Table 1.7: **Fallen angel buying strategy, counts and significance**. This table displays for the fallen angel buying strategy in Table 1.6 how many firms were included when using each of the utilized counterfactuals and also how many firms were common in both the synthetic control and in each of the other counterfactuals. This was done in order to assess if results were driven by only one or two extreme outliers and if there was any variation in the portfolio composition at all. Two-sided unpaired two-samples Wilcoxon tests were also conducted in order to asses whether the difference in mean total returns using the synthetic control or one of the other counterfactuals is statistically significant or not. *, **, ***, denotes significance on the 10, 5, 1 percent level, respectively.

\Counterfactual Metrics	Synthetic Control	Controls IG EW	Controls IG VW	Bloomberg IG VW
Panel A: Long (abnormal preannouncement return < 0) Holding period: 12 months				
# of firms	111	121	118	113
# of shared firms with synthetic control	–	104	100	95
p-value of two-sided unpaired two-samples Wilcoxon test	–	0.53	0.54	0.68
Panel B: Long (abnormal preannouncement return < 20th percentile of abnormal preannouncement return) Holding period: 12 months				
# of firms	35	35	35	35
# of shared firms with synthetic control	–	30	28	28
p-value of two-sided unpaired two-samples Wilcoxon test	–	0.75	0.57	0.52

Table 1.7 investigates for the fallen angel buying strategy in Table 1.6, how many firms were included when utilizing different counterfactuals. Further-

more, in order to test whether results are purely driven by one or two firms, the number of shared firms with the synthetic control are also displayed. Finally, in order to assess whether the found advantage of using the synthetic control in generating superior postannouncement returns is statistically significant vs. using a different counterfactual, a two-sided unpaired two-samples Wilcoxon test is conducted (the null hypothesis being, that the mean total returns when using the synthetic control vs. using a different counterfactual are equal). I opted for a non-parametric test, because a Shapiro–Wilk test on the total returns of the fallen angel buying strategy indicated non-normality. Panel A and Panel B in Table 1.7 cover the same setup as explained in Table 1.6. From Table 1.7 it can be seen, that there is variation among which firms end up in the portfolio and that the results are not just due to one extreme observation ending up in the portfolio. Figure 1.7 further corroborates this claim. It is, however, evident, that the outperformance of using the synthetic control as a counterfactual vs. a different choice of counterfactual, is not statistically significant. Nevertheless, the economic significance of the results can be substantial for investment professionals. In particular, when considering those fallen angels that experienced the most severe sell-off, an almost 6 percent advantage in using the synthetic control vs. the common industry Bloomberg benchmark index is material.

This research can help investment managers in constructing factor portfolios that systematically harvest this observed anomaly resulting from institutional and behavioral factors. In particular, investment managers need guidance on which fallen angels to buy and how long they should hold them in order to reap the vast amount of the fallen angel's price rebound. Various insights could be gained investigating the pre- and postannouncement returns of fallen angel bonds. Firstly, without considering their preannouncement sell-off size, one can outperform on average the high-yield market as a whole for 10 months after the downgrade announcement (cumulative outperformance of 4.41 percent after 10 months holding period). After that period, the outperformance is declining. Secondly, buying only fallen angel bonds trading at a discount vs. their counterfactual can increase total returns. In the case of using the synthetic control as a counterfactual, outperformance with respect to high-yield bonds increases to 8.06 percent for a holding period of 10 months. Only buying bonds that trade at a discount, therefore is able to almost double outperformance compared to the naive fallen angel buying strategy. Using the synthetic control and only buying bonds that trade at a discount further allows the strategy to outperform high-yield bonds up until a holding period of 12 months (albeit the bulk of the rebound has been recovered at month

10). Buying only fallen angels that suffered the most severe sell-off, further is able to increase total returns up to a value of 51.83 percent for a holding period of 12 months. Thirdly, using the synthetic control – albeit not being statistically significant – yields higher total returns than when using any of the other counterfactuals. When considering the case of the fallen angels that featured the biggest sell-off, the synthetic control has – for a holding period of 12 months – an average total return that is 5.77 percentage points higher than the one obtained from using the Bloomberg investment-grade index as a counterfactual.

Investment managers are therefore well advised to buy fallen angel bonds that trade at a discount. Whereas using a synthetic control would be best when wanting to assess the degree of overselling of the fallen angel bond – with respect to future investment gains after the downgrade – using any common counterfactual will yield a higher return than when simply buying fallen angel bonds without considering the size of their sell-off. Whereas the outperformance vs. high-yield bonds only continues up until 10 months after the downgrade for the naive fallen angel buying strategy, using the size of the sell-off allows this outperformance to continue up until 12 months.

1.7 Conclusion

A fallen angel bond is a corporate bond that was originally rated as investment-grade but got subsequently downgraded to speculative-grade by a rating agency. In this paper, I investigate the research question of how high the preannouncement sell-off of fallen angel bonds could be and how this could be used in making better informed investment decisions. The literature of the effect of bond reratings on bond prices relies heavily on event studies. Event studies analyze excess returns of fallen angel bonds with respect to some broad bond benchmark around the announcement date of the rerating by the rating agency. Due to their usage of a broad bond benchmark, the question arises whether the estimated effects can be viewed as adequate or not.

To overcome this issue that prior literature has faced, I propose a unified causal event study framework. This framework unifies both, the traditional event study framework and a commonly used framework from the causal inference literature, utilizing potential outcomes. Since the counterfactual of a fallen angel bond's return – i.e. the return that the fallen angel bond would have realized during the preannouncement sell-off had it not been downgraded

– is not observed, identifying assumptions must be made. Since event studies can be framed as involving only one treated unit (the fallen angel bond) and having a substantial number of controls (investment-grade bonds that did not experience a downgrade to speculative-grade), I propose to use a generalized version of the synthetic control method that can cope with these high-dimensional settings by applying a Lasso type penalty on the weights. Whether the estimated preannouncement sell-off of fallen angel bonds can be viewed as causal relies heavily on rather strict identifying assumptions. Due to the nature of financial markets with its abundance of peer-effects, externalities and spillover effects, as well as my definition of treatment that relies on the outcome 100 days later, these identifying assumptions probably are violated. However, since I am not per se interested in estimating the exact causal size of the preannouncement sell-off for each fallen angel bond, but rather using this estimate in order to make better informed buying decisions, I argue that the potential non-causality becomes less of an issue.

My findings are similar to those found in the past literature. In particular, I document a strong "V-curve" shape of the total returns around the downgrade date. This means that there exists an anticipation period where fund managers and investment professionals alike engage in a fire sale. In this paper I coin this period the preannouncement sell-off. After the downgrade announcement of the rating agency, a steep recovery sets in, suggesting that the price of the bond has been pushed well below its fair value. Reasons identified from the literature suggest that this is an anomaly driven by institutional and behavioral factors. On the one hand, fund managers who are constrained by their investment guidelines to not hold any speculative-grade assets – or insurance funds with similar constraints – must sell the fallen angel bonds immediately without regard to price or opportunity. The story with exchange traded funds is much the same, their objective is to track an index and objects that leave the index must be sold. Since investment professionals do not like to report downgrades of their portfolio or are constrained by some tracking error considerations, they are likely to sell-off these assets well in advance of the downgrade announcement, thus depressing the price of the security even before it becomes a fallen angel. In particular, I find that the preannouncement sell-off fallen angel bonds suffer amounts to -3.46 percent on average. Furthermore, the size of the preannouncement sell-off of fallen angels is strongly negatively related to future postannouncement performance. The relationship is more pronounced if only fallen angels are bought that trade at a discount. I was able to document, that – albeit not being statistically significant – using the synthetic control as a counterfactual, was able to outperform other common choices of counterfactuals. In particular if one focuses on the subset of fallen angels that experienced the most severe preannouncement sell-off, using a synthetic control generates a total return that is almost 6 percentage points higher than when using the industry standard Bloomberg benchmark bond index. Using this data-driven, innovative approach of finding a more representative counterfactual – a "novel" benchmark – can thus result in material economic gains for investment managers.

The impact of my work on the field of bond reratings and bond price effects is threefold. Firstly, by unifying the traditional event study framework with the framework from the causal inference literature, I am able to provide a fallen angel specific, custom benchmark return, rather than some vague broad-based benchmark return. Secondly, I propose the use of an already developed data-driven procedure which borrows from recent developments of the machine learning literature to impute the counterfactual, which leads in turn to the estimated preannouncement sell-off and its application in making better informed fallen angel buying decisions after the downgrade announcement. Thirdly, I demonstrate the new methodology on a state-of-the-art dataset that features prices from the OTC-market with daily frequency. In comparison to prior literature, my fallen angels sample is quite rich with 176 unique firm-level observations. Bolognesi et al. (2014) report 48 and May (2010) 105 fallen angel firm-level bonds. With my empirical study, I am able to further corroborate the "V-shape" curve of total returns around the downgrade announcement already identified in the literature. The impact on a potential fallen angel factor portfolio construction is therefore huge, where questions of when to exit and how to weight individual trades arise.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, *59*(2), 495–510.

Abadie, A., & Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *The American Economic Review*, *93*(1), 113–132.

Ambrose, B. W., Cai, K. N., & Helwege, J. (2012). Fallen Angels and Price Pressure. *The Journal of Fixed Income*, *21*(3), 74–86.

Ambrose, B. W., Cai, N. K., & Helwege, J. (2008). Forced Selling of Fallen Angels. *The Journal of Fixed Income*, *18*(1), 72–85.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Athey, S., & Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32.

Bai, J., Bali, T. G., & Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, *131*(3), 619–642.

Bessembinder, H., Kahle, K. M., Maxwell, W. F., & Xu, D. (2009). Measuring Abnormal Bond Performance. *The Review of Financial Studies*, *22*(10), 4219–4258.

Bolognesi, E., Ferro, M., & Zuccheri, A. (2014). The Impact of Fallen Angels on Investment Grade Corporate Bonds Portfolios: Evidence from the European Market. *International Journal of Finance & Economics*, *19*(4), 267–278.

Brown, S. J., & Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics*, *14*(1), 3–31.

Burman, P., Chow, E., & Nolan, D. (1994). A Cross-Validatory Method for Dependent Data. *Biometrika*, *81*(2), 351–358.

Cantor, R., Gwilym, O. A., & Thomas, S. H. (2007). The Use of Credit Ratings in Investment Management in the U.S. and Europe. *The Journal of Fixed Income*, *17*(2), 13–26.

Chernozhukov, V., Wüthrich, K., & Zhu, Y. (2021). An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls. *Journal of the American Statistical Association*.

Clare, A., Thomas, S., & Motson, N. (2016). *Fallen Angels: The Investment Opportunity* (Working Paper). Cass Business School.

Cox, D. R. (1958). *Planning of Experiments.* Wiley.

De Bondt, W. F. M. D., & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance*, *40*(3), 793–805.

Dick-Nielsen, J. (2009). Liquidity Biases in TRACE. *The Journal of Fixed Income*, *19*(2), 43–55.

Dick-Nielsen, J. (2014). *How to Clean Enhanced TRACE Data* (SSRN Working Paper No. 2337908).

Dor, A. B., & Xu, Z. (2011). Fallen Angels: Characteristics, Performance, and Implications for Investors. *The Journal of Fixed Income*, *20*(4), 33–58.

Doudchenko, N., & Imbens, G. W. (2017). *Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis* (arXiv Working Paper No. 1610.07748).

Ellul, A., Jotikasthira, C., & Lundblad, C. T. (2011). Regulatory pressure and fire sales in the corporate bond market. *Journal of Financial Economics*, *101*(3), 596–620.

Emery, K., & Gates, D. (2014). *Why Fallen Angels Fall: An Examination of Nonfinancial Corporate Fallen Angels 1999-2013 and 2014 Outlook* (Special Comment No. 162807). New York, NY: Moody's Investors Service.

Ferman, B., & Pinto, C. (2019). Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity. *The Review of Economics and Statistics*, *101*(3), 452–467.

Firpo, S., & Possebom, V. (2018). Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets. *Journal of Causal Inference*, *6*(2).

Grier, P., & Katz, S. (1976). The Differential Effects of Bond Rating Changes Among Industrial and Public Utility Bonds by Maturity. *The Journal of Business*, *49*(2), 226–239.

Hand, J. R. M., Holthausen, R. W., & Leftwich, R. W. (1992). The Effect of Bond Rating Agency Announcements on Bond and Stock Prices. *The Journal of Finance*, *47*(2), 733–752.

Heckman, J. J., Lalonde, R. J., & Smith, J. A. (1999). The Economics and Econometrics of Active Labor Market Programs. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3, pp. 1865–2097). Elsevier.

Hite, G., & Warga, A. (1997). The Effect of Bond-Rating Changes on Bond Price Performance. *Financial Analysts Journal*, *53*(3), 35–51.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In A. Tversky, D. Kahneman, & P. Slovic (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 414–421). Cambridge: Cambridge University Press.

Kaul, A., Klößner, S., Pfeifer, G., & Schieler, M. (2021). Standard Synthetic Control Methods: The Case Of Using All Preintervention Outcomes Together With Covariates. *Journal of Business & Economic Statistics*.

Kinn, D. (2018). *Synthetic Control Methods and Big Data* (arXiv Working

Paper No. 1803.00096).

Li, K. T. (2020). Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods. *Journal of the American Statistical Association*, *115*(532), 2068–2083.

May, A. D. (2010). The impact of bond rating changes on corporate bond prices: New evidence from the over-the-counter market. *Journal of Banking & Finance*, *34*(11), 2822–2836.

O'Neill, S., Kreif, N., Grieve, R., Sutton, M., & Sekhon, J. S. (2016). Estimating causal effects: Considering three alternatives to difference-in-differences estimation. *Health Services & Outcomes Research Methodology*, *16*, 1–21.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: Hv-block cross-validation. *Journal of Econometrics*, *99*(1), 39–61.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Rubin, D. B. (1980). Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association*, *75*(371), 591–593.

Steiner, M., & Heinke, V. G. (2001). Event study concerning international bond price effects of credit rating actions. *International Journal of Finance & Economics*, *6*(2), 139–157.

Wansley, J. W., & Clauretie, T. M. (1985). The Impact of Creditwatch Placement on Equity Returns and Bond Prices. *Journal of Financial Research*, *8*(1), 31–42.

Wansley, J. W., Glascock, J. L., & Clauretie, T. M. (1992). Institutional Bond Pricing and Information Arrival: The Case of Bond Rating Changes. *Journal of Business Finance & Accounting*, *19*(5), 733–750.

Weinstein, M. I. (1977). The effect of a rating change announcement on bond price. *Journal of Financial Economics*, *5*(3), 329–350.

Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Political Analysis*, *25*(1), 57–76.

1.A Fallen angel and counterfactual data

Figure 1.8: **Mean fallen angel and counterfactual return until event**. This figure displays averaged total returns for fallen angels and its 4 counterfactuals: 1) synthetic control, 2) controls equally-weighted, 3) controls value-weighted and 4) Bloomberg investment-grade value-weighted index.

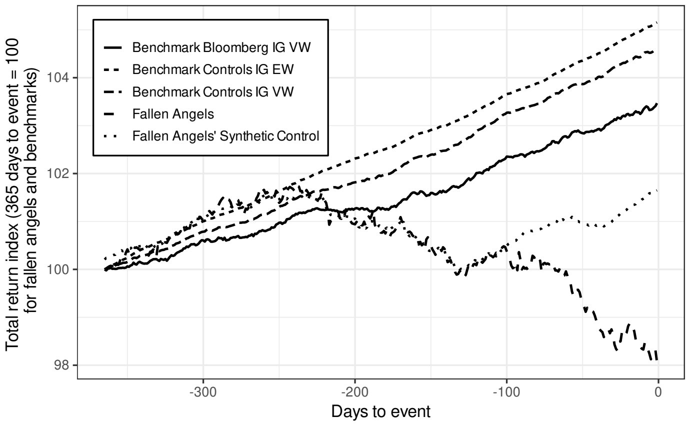
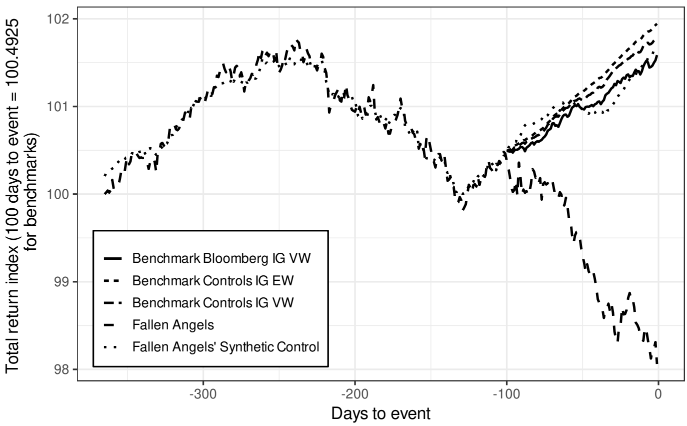


Figure 1.9: **Mean fallen angel and counterfactual return until event (indexed at 100 days to event)**. This figure displays averaged total returns for fallen angels and its 4 counterfactuals: 1) synthetic control, 2) controls equally-weighted, 3) controls value-weighted and 4) Bloomberg investment-grade value-weighted index.



Chapter 2

The cross-section of corporate bond returns

Co-author: Dimitrios Nteventzis

Abstract

We study the cross-section of corporate bonds utilizing a large set of financial statements, equity and bond characteristics. We use a predictive regression framework and the adaptive Lasso to choose the most relevant characteristics for the cross-section of corporate bonds. Applying the adaptive Lasso to the full dataset, we find a ten-factor model, with value, bond reversal, and equity momentum spillover being the dominant factors. Contrary to equity studies, financial variables from Compustat do not appear to have strong power in predicting corporate bond returns. We validate our initial results by running an out-of-sample exercise using an expanding window approach. Out of the 60 months utilized in the out-of-sample, the adaptive Lasso consistently chooses value, bond reversal, and equity momentum spillover. Finally, we evaluate the economic benefits of investing according to the predictions of the adaptive Lasso and find significant benefits in terms of absolute and risk-adjusted returns.

2.1 Introduction

Most of the academic research after the milestone publication of Fama and French (1992) has focused on discovering factors in the equity space. Despite the increasing size of the corporate debt markets and its importance as a financing channel for firms, the cross-section of corporate bond returns remained under-researched. However, a rapidly expanding literature has recently focused on asset-pricing factors in the corporate bond market. The improved data availability on over-the-counter (OTC) bond transactions and the need to jointly study returns of different asset classes to identify significant factors of the stochastic discount factor (SDF) are critical drivers for the increased attention on corporate bonds.

The purpose of this paper is to explore the cross-section of corporate bond returns. Similar to studies focusing on equities, one of the biggest challenges in explaining corporate bond returns is the large number of potential characteristics that affect the cross-section. To overcome this issue, we employ the adaptive Lasso in order to identify significant characteristics for the cross-section of corporate bonds. We utilize a large dataset of characteristics which is comprised of 46 equity characteristics similar to Freyberger, Neuhierl, and Weber (2020) and 21 bond specific characteristics. We separate our paper into two parts. In the first part, we offer a detailed description of the bond data and characteristics, and we perform an initial analysis by forming bi-variate sorted portfolios. In the second part, we use the adaptive Lasso in a pooled predictive regression setting in order to choose significant characteristics. Similar to portfolio sorts, our analysis focuses explicitly on the cross-sectional dimension of expected returns. To measure the economic benefits of the proposed model, we create a single factor by sorting bonds each month according to the 1-step ahead forecasted return. We then test the significance of the factor against a variety of previously reported models.

First, our initial analysis of forming factors through sorted portfolios can be interpreted as a detailed replication study. As demonstrated by Hou, Xue, and Zhang (2020) most reported anomalies in the equity space fail to replicate even before accounting for the issue of multiple testing bias that was brought to attention by Harvey, Liu, and Zhu (2016). Given the lower quality of available data for bonds vs. equity and the idiosyncrasies of bonds (defaults, fixed maturity/discontinuity), we argue that replication studies are much more needed in corporate bonds than in equities. Many previous studies merge data from the Trade Reporting and Compliance Engine (TRACE) with

Thomson Reuters Datastream and Lehman Brothers fixed income databases to increase the sample size (TRACE starts in 2002). To increase the reliability of our study, we employ only trade-based data from TRACE and not dealer quotes reported in Datastream and Lehman Brothers databases. We provide an extensive list of the steps we take to clean the TRACE data and arrive at our final sample. Contrary to previous studies, we directly control for defaulted bonds by matching our dataset to the Altman[1] default database. From our initial analysis, we find that the dominant factors are the value factor proposed by Israel, Palhares, and Richardson (2018) and the one-month bond reversal and equity momentum spill-over reported by Chordia, Goyal, Nozawa, Subrahmanyam, and Tong (2017) and Gebhardt, Hvidkjaer, and Swaminathan (2005), respectively. Interestingly, we find that downside risk factors proposed by Bai, Bali, and Wen (2019), the momentum factor Jostova, Nikolova, Philipov, and Stahel (2013) and illiquidity factor Bao, Pan, and Wang (2011) do not replicate in our sample. Our hypothesis is that downside and illiquidity risk factors are sensitive to the treatment of defaulted bonds.

Second, our paper contributes to the study of the cross-section of corporate bonds. Previous papers have focused on a handful of pre-selected equity or bond factors to explain corporate bonds' returns. We extend previous work by using a large set of characteristics and utilizing modern machine learning techniques in empirical asset pricing similar to Gu, Kelly, and Xiu (2020). Using the adaptive Lasso, we choose a ten-factor model with the value factor proposed by Israel et al. (2018), the one-month bond reversal and equity momentum spill-over being the most important. Contrary to equity studies, financial variables from Compustat do not appear to have strong power in predicting corporate bond returns. To validate our results and test the usefulness of adaptive Lasso, we run an out-of-sample exercise, using an expanding window approach and forecasting the returns of the whole cross-section each month. We then construct a factor portfolio (F^{lasso}) by going long/short the bonds with forecasted returns in the top/bottom 20%. The adaptive Lasso factor generates significant monthly excess returns of 0.28%. Using intercept tests, we show that F^{lasso} carries a positive and significant premium which can not be explained by existing models. The risk-adjusted excess return of the factor, expressed by the intercepts, is significant and relatively stable, ranging from 0.24% to 0.32% even though we test for a large set of competing models. We thus conclude that the adaptive Lasso framework can deliver superior returns for an investor who invests according to the forecasted returns of the model.

[1] Altman-Kuehne NYU Salomon Center Corporate Bond Default Master Database.

Our paper is organized as follows. In the second section, we provide a detailed literature review of reported factors in corporate bonds and machine learning techniques employed in empirical asset pricing. In the third section, we present our data. In the fourth section, we perform an initial analysis by forming bi-variate sorted portfolios according to all 67 characteristics that we study. In section five, we use the adaptive Lasso to choose characteristics and evaluate the economic benefits of the proposed model over previously reported models. Section six concludes our findings.

2.2 Literature review

The existing literature on corporate bond returns can be separated into two broad categories: The first using equity and financial characteristics and the second using bond-specific characteristics. In the first category, researchers exploit the fact that the fundamentals of the same firm drive both equity and corporate debt returns. Thus, they use well-established factors from the equity asset-pricing literature to price the cross-section of corporate bonds. Chordia et al. (2017) try to answer whether market anomalies are common in the equity and corporate bond market. They use ten well-established equity characteristics and conclude that equity factors are weak in the corporate bond universe.[2] Choi and Kim (2018) use six equity characteristics to study the market integration between equity and corporate debt.[3] They also find a weak presence of equity factors in the corporate bond market. Finally, Bektić, Wenzler, Wegener, Schiereck, and Spielmann (2019) use the four Fama-French factors[4] to price U.S. and European corporate bonds. They get economically significant results for the U.S. high yield market but weak results for the U.S. and European investment grade market. Results among those papers are contradicting. Bektić et al. (2019) find significant results and the same signs as the equity studies for U.S. high yield. Chordia et al. (2017) report significant results for the profitability factor but an opposite sign compared to the equity studies for the value factor. Finally, Choi and Kim (2018) report not significant results for both value and profitability.

[2] Chordia et al. (2017) use: Size, value, momentum, reversal, accruals, asset growth, profitability, net issuance, earnings surprise, and idiosyncratic volatility.

[3] Choi and Kim (2018) use: Asset growth, investment, profitability, net issuance, value and momentum.

[4] Small minus big (SMB), high minus low (HML), robust minus weak (RMW), and conservative minus aggressive (CMA).

The weak performance of equity-specific factors in pricing the cross-section of corporate bonds highlights the differences between the two markets and the need to construct factors modified to price the cross-section of corporate bonds. Correia, Richardson, and Tuna (2012) are the first to report a value factor in corporate bonds. Instead of using book-to-market, they construct the value factor by comparing market and model implied spreads estimated from a structural credit risk model *á la* Merton (1974). Jostova et al. (2013) report strong evidence of momentum effects in the U.S. corporate bond market, mostly driven by the high yield segment. Israel et al. (2018) and Houweling and van Zundert (2017), propose two similar four-factor models that include value, momentum, quality/low-risk and carry/size. Finally, Bai et al. (2019) introduce downside risk as a new factor in corporate bonds and propose a four-factor model that includes the market, downside risk, credit risk, and liquidity.

Our paper is also related to the expanding literature on machine learning and asset pricing. Gu et al. (2020) perform a comparative analysis of machine learning methods in choosing firm characteristics. They find significant economic benefits over the standard ordinary least squares (OLS) estimation method. All machine learning methods agree that the dominant set of characteristics includes variations of momentum, liquidity, and volatility. Feng, Giglio, and Xiu (2020) propose a regularized two-pass estimation procedure focusing on risk-prices instead of risk-premia and treating for the effect of omitted variables. They report that chosen factors vary from 10 in 1994 to 18 in 2016. They also find a significant effect for some recently suggested factors, noting that academic research continues to discover new factors. Messmer and Audrino (2017) apply the adaptive Lasso methodology to a total of 64 firm characteristics and find 14 of them being significant. Finally, Freyberger et al. (2020) highlight the importance of nonlinearities in the relationship of characteristics and expected returns and apply the grouped Lasso methodology in order to identify useful characteristics non-parametrically. They find that 13 variables have incremental explanatory power for expected returns.

2.3 Data and variable definition

2.3.1 Corporate bond data

We use OTC bond transaction data which is available directly through Financial Industry Regulatory Authority's (FINRA) Trade Reporting and Compliance Engine (TRACE). TRACE was introduced by the National Association of Securities Dealers (NASD) in 2002 to improve transparency in the OTC corporate bond market. From February 2005 onwards, 99 percent of all TRACE-eligible bond transactions are covered on an intra-day basis. The information covered in the TRACE database is listed by transaction, and key variables include transaction date, time, price, and traded volume. Therefore, TRACE is the most comprehensive source of pricing information when it comes to research questions concerning the U.S. corporate bond market. We use the enhanced version of the TRACE database, which has no volume cap on reported trades and thus captures a broader range of the U.S. corporate bond market. Our data extend from July 2002 to December 2018. Since the TRACE data is dealer-reported, errors can occur. Instead of correcting the data in the database directly, trade messages are appended, indicating either cancellations, corrections, or reversals. Trades can also be double-counted in the TRACE system because various parties can report the same trade. We follow the steps outlined in Dick-Nielsen (2009) and Dick-Nielsen (2014) to take care of the cancellation, correction, reversal, and double counting issues.

In our effort to clean the TRACE data, we follow the cleaning steps outlined in Bessembinder, Kahle, Maxwell, and Xu (2009) and Bai et al. (2019). In particular, we remove transaction records: (1) with trading volume of less than 10,000 USD, (2) are labeled as when-issued, (3) locked-in, (4) have special sales conditions, (5) have more than a two-day settlement, and (6) are flagged as equity-linked notes. In order to minimize the effect of bid-ask spreads in prices, we calculate the daily clean price as the trading volume-weighted average of intra-day prices. We continue by following Bai et al. (2019) in removing trade records that feature a transaction price under 5 or above 1,000 USD. This step implicitly removes some defaulted bonds. However, this is a very low threshold in our view since many defaulted bonds tend to trade above 5 USD. Including defaulted bonds in our study can create biased results for three main reasons: (1) defaulted bonds typically do not accrue interest, (2) liquidity is typically very low after a default, and (3) actual recovery is deal-specific and hard to estimate. For these reasons and in contrast to previous papers such as Bai et al. (2019), we choose to control for defaulted bonds directly. We incorporate the issue-level default data information from the Altman[5] database and exclude all future bond observations post a given default date. In the initial results, we observed an overly strong reversal factor. After checking data manually, we found that extreme day-to-day reversals exist due to prices of small trades and thus is a type of an outlier. For this reason, we implemented a reversal rule, where we exclude trade records that featured a daily reversal of more than 10 USD in absolute terms. E.g., if a bond trades at 100 on Monday, drops to below 90 on Tuesday, and then features again a price of over 100 on Wednesday, we exclude the trade record for Tuesday for that particular bond. We then aggregate the daily bond price data to the monthly frequency. We follow Bai et al. (2019) in only considering a bond's monthly end price as a non-missing value if it falls within the last seven weekdays of that particular month. The monthly traded volume for each bond is generated by first aggregating it to the daily level by taking the intra-day sum for all days and then taking the sum over all days in a particular month.

We then match the bond CRSP link database to get the corresponding firm identifier for each bond and match the corresponding equity and financial variables from CRSP and Compustat. We retrieve bond issue information such as the maturity date, coupon rate, coupon payment frequency, etc., from the Mergent FISD database. A key metric of a bond is its rating and how it changes over time. We access the historical rating history of each bond through Mergent FISD's historical bond rating database. It includes ratings from Standard & Poor's, Moody's, and Fitch. We follow the industry convention and form composite ratings by applying the following methodology: if three outstanding ratings are available, we choose the middle rating; if two ratings are outstanding, we choose the more conservative one, and if only one rating is outstanding, we choose the one that is available.

We use information from the Mergent FISD database in order to further filter our data. We exclude bonds that are not listed or traded in the U.S. public market. That includes private placements, 144A bonds, bonds that do not trade in U.S. dollars, and issuers not located in the U.S. We further remove structured notes, mortgage-backed securities, asset-backed securities, agency-backed bonds, and convertible bonds. We only keep bonds with a fixed or a zero-coupon and exclude all bonds with a variable or floating coupon rate. In addition to excluding perpetuals (bonds without a fixed maturity), we exclude bonds with a remaining lifetime of more than 30 years since these bonds tend to be illiquid. If a bond trades close to maturity, i.e. less than one year, it

[5] Altman Kuehne NYU Salomon Center Corporate Bond Default Master database.

is delisted from major corporate bond indices, thus reducing its liquidity. We thus remove bonds with less than one year remaining to maturity.

Afterwards, we calculate for each bond-month observation its corresponding yield-to-maturity and modified duration. Both of these variables are key metrics needed in the construction of characteristics. In order to calculate the total return of a bond, we also need to incorporate information on when coupon payments were made and what the accrued interest amounted to at each point in time for each individual bond issue. The coupon payment schedule can be backed out by going backwards from the last coupon payment by the bond's coupon payment frequency. Accrued interest is the amount of interest that accrues from one coupon payment to the next. We follow Bessembinder et al. (2009) in their calculation of a bond's total return:

$$r_{i,t} = \frac{P_{i,t} + AI_{i,t} + C_{i,t}}{P_{i,t-1} + AI_{i,t-1}} - 1, \tag{2.1}$$

where $P_{i,t}$ is the transaction price (i.e. the clean price), $AI_{i,t}$ the accrued interest and $C_{i,t}$ the coupon payment if any, of bond i in month t. We calculate excess returns using the maturity-matched treasury returns instead of the one month risk-free rate. We denote bond i's excess return as $R_{i,t} = r_{i,t} - r_{f,t}$, where $r_{f,t}$ is the return of bond i's maturity-matched treasury bond. Available constant maturity treasury bonds cover maturities including 1, 2, 5, 7, 10, 20 and 30 years. As a final step in our sample setup, we adopt the same bond trading/liquidity restriction of Bai et al. (2019). In particular, for a bond-month observation to be considered in our analysis, we require that it had a valid return for at least 24 out of the past 36 months.

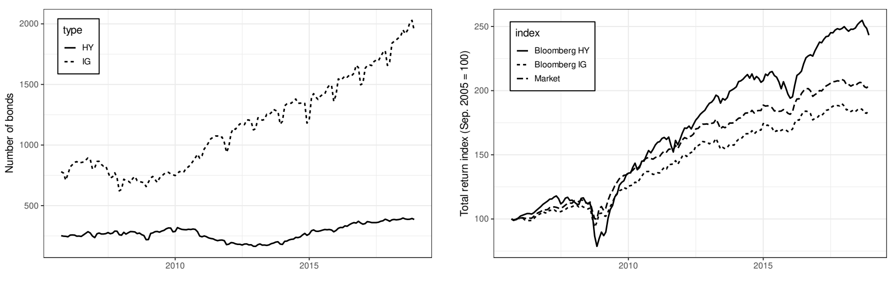
Table 2.1 describes the steps we take to filter the TRACE dataset and the impact of those steps on the dimensions of our dataset. The sample starts with the intra-day level frequency of the raw TRACE data consisting of 184,645 bond issues and ends with a final sample size of 7,839 bond issues from 1,087 unique firms. The largest loss of observations is when daily observations are aggregated to monthly. This step reduces the number of issues by roughly 60 percent. This reduction is mainly due to the restriction that to have a non-missing return in a given month, the bond needs to trade in the last seven weekdays of two consecutive months. After matching the Mergent FISD database and applying the filtering steps described above, the number of issues reduces from 48,598 to 25,994. A sizeable fraction of observations is lost due to the construction of bond and equity/financial variables. The proportionally bigger loss in observations when matching in the bond variables is

Table 2.1: **Sample setup**. In this table we describe the detailed steps we take to filter the TRACE dataset and the impact of those steps on the dimensions of our dataset.

	issues	firms	frequency	obs
starting sample, TRACE raw data	184,645		intra-day	223,961,512
- daily aggregation	173,017		daily	22,063,238
- remove trade records, price < 5, > 1000 USD	172,422		daily	22,024,034
- remove trade records, price reversal > |10| USD	172,422		daily	22,005,801
- monthly aggregation	66,026		monthly	1,594,457
- match bond CRSP link database	48,618	3,022	monthly	1,152,419
- match bond default database	48,598	3,015	monthly	1,142,465
- match Mergent FISD issue database	25,994	1,889	monthly	738,947
- match Mergent FISD ratings database	24,002	1,775	monthly	696,258
- remove trade records, < 24 out of 36 months traded	22,955	1,744	monthly	619,498
- match bond variables	10,836	1,254	monthly	336,707
- match equity and financial variables	7,839	1,087	monthly	227,795

due to the necessity of a 36 month rolling window in calculating specific bond characteristics such as the value at risk or expected shortfall metrics from Bai et al. (2019). Considering equity and financial variables further reduces the sample size because some firms are not listed or do not file financial statements.

Figure 2.1: **Key characteristics of the U.S. corporate bond market**. In the left panel we show the time evolution of the number of bonds split by IG and HY. In the right panel we show the cumulative return of Bloomberg IG/HY benchmark indexes and the market created from our sample. Data from October-2005 to December-2018.



(a) Corporate bond market development (b) Corporate bond market development

Figure 2.1a shows the number of bonds in our estimation sample, split by rating category. The sample is predominantly investment grade (IG) rated, with the proportion of investment grade increasing since 2010. The amount of high yield (HY) bonds in our sample is reasonably constant over time, with an average of roughly 250 bonds each month. Figure 2.1b shows the cumulative performance of the Bloomberg IG/HY index and the value-weighted market from our sample. Since we include both investment grade and high yield bonds in our sample, we expect the market return to track the investment grade benchmark closer but have a higher return due to the inclusion of high yield bonds. The investment grade and high yield Bloomberg benchmark indices have an annualized mean (standard deviation) of 4.81 (5.49) and 7.18 (9.45) percent, respectively. The market constructed from our sample has an annualized mean (standard deviation) of 5.52 (5.12) percent.

Table 2.2 shows the summary statistics for our sample of corporate bonds. The mean, median, standard deviation and percentiles were calculated by pooling all bond-month observations. On average, corporate bonds exhibited a total return of 0.42 percent in a given month. 81 percent of observations can be categorized as investment grade. Bonds on average were issued 6.52 years ago and have a remaining 8.53 years until maturity. The average coupon and yield to maturity are 5.70 and 4.64 percent,while the average amount issued is 650.37 million USD.

For the asset pricing part where we conduct intercept tests, we need to control for the default (DEF) and the term (TERM) factor. We follow Bai et al. (2019) in the construction of these two factors. The default factor is defined as the difference between the equal-weighted market portfolio return of corporate bonds and the 30-year government bond return. The term factor is defined as the difference between the monthly 30-year government bond return and the return on holding the one-month Treasury bill.

Table 2.2: **Summary statistics (CRSP/Compustat matched)**. In this table we display the summary statistics of our main sample which is CRSP/Compustat matched. Data from October-2005 to December-2018.

					Percentiles					
	N	Mean	Median	SD	1st	5th	25th	75th	95th	99th
Bond total return (percent)	227,795	0.42	0.31	3.70	-7.95	-2.89	-0.34	1.16	3.81	8.82
Rating (1-21, 1=AAA, IG $\leq$ 10)	227,795	8.27	8	3.44	1	3	6	10	15	18
Investment grade (1=IG, 0=HY)	227,795	0.81	-	-	-	-	-	-	-	-
Time to maturity (years)	227,795	8.53	5.25	7.93	1.08	1.42	3.00	11.58	25.58	26.83
Age (years)	227,795	6.52	5.42	3.81	3.08	3.17	3.92	7.58	15.17	21.42
Coupon (percent)	227,795	5.70	5.75	1.62	1.75	2.88	4.75	6.75	8.25	9.50
Yield to maturity (percent)	227,795	4.64	4.17	4.52	0.67	1.22	2.64	5.63	8.88	17.79
Modified duration (years)	227,795	5.77	4.47	4.08	1.08	1.34	2.71	7.98	13.98	15.71
Offering amount (million USD)	227,795	650.37	500.00	646.70	17.04	30.75	300.00	750.00	2000.00	3000.00

Table 2.3: **Summary statistics (Not CRSP/Compustat matched)**. In this table we display the summary statistics of our sample when we do not match with the CRSP/Compustat databases. Data from October-2005 to December-2018.

					Percentiles					
	N	Mean	Median	SD	1st	5th	25th	75th	95th	99th
Bond total return (percent)	336,707	0.54	0.34	6.00	-11.78	-3.46	-0.38	1.30	4.58	12.52
Rating (1-21, 1=AAA, IG $\leq$ 10)	336,707	8.86	8	3.77	1	4	6	10	16	20
Investment grade (1=IG, 0=HY)	336,707	0.75	-	-	-	-	-	-	-	-
Time to maturity (years)	336,707	8.81	5.50	7.88	1.08	1.42	3.17	13.25	25.50	26.83
Age (years)	336,707	6.84	5.58	4.05	3.08	3.25	4.08	7.92	16.17	21.83
Coupon (percent)	336,707	5.91	6.00	1.64	1.85	3.10	5.00	6.95	8.50	9.88
Yield to maturity (percent)	336,707	5.46	4.51	7.38	0.71	1.32	2.85	6.16	10.99	29.49
Modified duration (years)	336,707	5.83	4.60	3.98	1.08	1.36	2.79	8.38	13.72	15.52
Offering amount (million USD)	336,707	619.50	500.00	618.78	17.28	30.00	250.00	750.00	1998.33	3000.00

2.3.2 Bond, equity and financial characteristics

As far as the characteristics for the machine learning step are concerned, Freyberger et al. (2020) compiled a comprehensive list of characteristics that potentially provide incremental information for the cross-section of expected returns. Due to their analysis being exclusively for equities, we also include relevant bond market characteristics that are used in the literature. A comprehensive overview of all 67 characteristics used in this paper is given in Table 2.4, while a detailed description of how each variable is constructed can be found in Appendix 2.C. The characteristics are grouped into the following categories: (1) past returns, (2) investment, (3) profitability, (4) intangibles, (5) value, and (6) trading frictions. In contrast to Freyberger et al. (2020), we use quarterly information whenever possible instead of yearly. We use the release date instead of the reporting date in order to avoid any forward-looking bias.

For bond characteristics we use age, life, modified duration, offering amount and rating. We include downside risk variables calculated from bond returns as in Bai et al. (2019) and past bond returns for different momentum variables. We also include the illiqudity variable of Bao et al. (2011) and value variables constructed only from bond data as in Israel et al. (2018) and Houweling and van Zundert (2017). In terms of the CRSP/Compustat data,[6] we use accruals, operating leverage and tangibility as defined by Hahn and Lee (2009) for intangibles. For profitability, we use a total of sixteen variables, with well-known performance indicators such as return on equity (ROE) and return on assets (ROA) among them. For trade frictions we use a total of four different

[6]Data from CRSP covers equity related data such as the stock price, trading volume, etc., whereas data from Compustat covers balance sheet and income statement information.

liquidity proxies for equities. For value we use a total of thirteen variables with well established measures such as book-to-market and price-to-earnings among them. Finally, we use equity past returns for different equity momentum indicators.

2.4 Initial analysis: Portfolio sorts

In this part, we perform a preliminary analysis on the impact of the characteristics on corporate bond returns. The purpose is twofold: Firstly, we are interested in which factors appear to be significant and secondly, we want to see which factors fail to replicate. For our initial analysis we rely on portfolio sorts which is the workhorse method in traditional asset pricing. In his presidential address Cochrane (2011) mentions that: "Looking at portfolio average returns rather than forecasting regressions was really the key to understanding the economic importance of many effects."

Portfolio sorts are the most common way of mapping characteristics to returns. The procedure is the following; for every month t, excess returns are sorted in quantile portfolios based on a specific characteristic's cross-sectional rank. Then, the factor is constructed by taking the difference between the returns of the two extreme portfolios. We follow Bai et al. (2019) and create traded factors from bi-variate sorts where we use credit rating as the first sorting variable, resulting in 25 (5x5) portfolios. Credit risk is a crucial driver of corporate bond returns; using the rating as a first sorting variable allows us to create portfolios with similar credit risk profiles. We construct each factor as the value-weighted average return difference of the extreme quantile portfolios among all credit rating portfolios.

To briefly illustrate how we construct each factor, consider that we have $R_{i,t}$ excess returns of N test assets, and we want to construct a factor for a single characteristic $C_{i,t}$. Then we denote $P_t^{1,5}$ the return at time t of a value-weighted portfolio of assets that belongs in the 1st quantile concerning C_{it} and the 5th quantile concerning credit rating. Then the factor is defined as:

$$\text{Factor}_t = \frac{1}{5}\sum_{j=1}^{5} P_t^{5,j} - \frac{1}{5}\sum_{j=1}^{5} P_t^{1,j} \tag{2.2}$$

In Table 2.5 we summarize the results from our portfolio sorts. We report

Table 2.4: **Bond, equity and financial characteristics**

(1)	β^b	Bond beta to the market, last 36 months
	Bond characteristics:	
(2)	Age	Months since the bond was issued
(3)	Life	Remaining months to maturity
(4)	Mdur	Modified duration
(5)	Offer-amt	Offering amount of the bond
(6)	Rating	Numeric value corresponding to rating
	Downside risk	
(7)	ES^b_{10}	10% exp. shortfall of each bond, last 36 months
(8)	ES^b_5	5% exp. shortfall of each bond, last 36 months
(9)	VaR10	10% VaR of each bond, last 36 months
(10)	VaR5	5% VaR of each bond, last 36 months
	Intangibles	
(11)	AOA	Absolute value of operating accruals
(13)	OA	Operating accruals
(13)	OL	COGS + SG&A to total assets (AT)
(14)	TAN	Tangibility
	Investment	
(15)	ΔCEQ	% Change in book equity
(16)	ΔPI2A	Change in PP&E and inventory over lagged AT
(17)	ΔShout	% Change in shares outstanding
(18)	INV	% Change in AT
(19)	IVC	% Change in inventory over AT
	Past Returns:	
(19)	Sprd^6	Cumulative spread change, $t-6$ to $t-1$
(20)	r^b_{1-0}	Bond return $t-1$
(21)	r^b_{2-1}	Bond return $t-2$ to $t-1$
(22)	r^b_{6-2}	Cumulative bond return, $t-6$ to $t-2$
(23)	r^b_{12-2}	Cumulative bond return, $t-12$ to $t-2$
(24)	r^b_{12-7}	Cumulative bond return, $t-12$ to $t-7$
(25)	r^b_{36-13}	Cumulative bond return, $t-36$ to $t-13$
(26)	r^e_{1-0}	Equity return $t-1$
(27)	r^e_{2-1}	Equity return $t-2$ to $t-1$
(28)	r^e_{6-2}	Cumulative equity return, $t-6$ to $t-2$
(29)	r^e_{12-2}	Cumulative equity return, $t-12$ to $t-2$
(30)	r^e_{12-7}	Cumulative equity return, $t-12$ to $t-7$
(31)	r^e_{36-13}	Cumulative equity return, $t-36$ to $t-13$
	Profitability:	
(32)	ATO	Sales to lagged net operating assets
(33)	CTO	Sales to lagged AT
(34)	Δ(ΔGM − ΔS)	Δ(% change in gross margin and % change in sales)
(35)	EPS	Earnings per share
(36)	IPM	Pretax income over sales
(37)	NOA	Net operating assets
(38)	PCM	Sales minus COGS to sales
(39)	PM	OI after depreciation over sales
(40)	Prof	Gross profitability over BE
(41)	RNA	OI after depreciation to lagged net operating assets
(42)	ROA	Income before extraordinary items to lagged AT
(43)	ROC	Size + long-term debt - AT to cash
(44)	ROE	Income before extraordinary items to lagged BE
(45)	ROIC	Return on invested capital
(46)	S2C	Sales to cash
(47)	SAT	Sales to AT
	Trade frictions	
(48)	AT	Total assets
(49)	Trade^b_{vol}	Monthly trading volume of the bond
(50)	Illiq	Illiquidity of the bond as in Bao et al. (2011)
(51)	LME	Price times shares outstanding
(52)	L_{turn}	Last month's volume to shares outstanding
	Value	
(53)	A2ME	AT to size
(54)	BEME	Book to market ratio
(55)	C	Cash to AT
(56)	C2D	Cash flow to total liabilities
(57)	δSO	Log change in split-adjusted shares outstanding
(58)	D2P	Total debt to size
(59)	E2P	Income before extraordinary items to size
(60)	LDP	Trailing 12-months dividends to price
(61)	NOP	Net payouts to size
(62)	O2P	Operating payouts to market capitalization
(63)	S2P	Sales to price
(64)	$Sales_g$	Sales growth
(65)	Q	Tobin's Q
(66)	Value^{ips}	Value as in Israel et al. (2018)
(67)	Value^{hvz}	Value as in Houweling and van Zundert (2017)

the mean of each factor and also the intercept from regressing each respective factor on three baseline models:

$$\text{Factor}_t = \alpha_1 + \beta_1^{mkt}\text{MKT}_t + \epsilon_{1,t} \; , \tag{2.3}$$

$$\text{Factor}_t = \alpha_2 + \beta_2^{mkt}\text{MKT}_t + \beta_2^{term}\text{TERM}_t + \epsilon_{2,t} \; , \tag{2.4}$$

$$\text{Factor}_t = \alpha_3 + \beta_3^{mkt}\text{MKT}_t + \beta_3^{term}\text{TERM}_t + \beta_3^{def}\text{DEF}_t + \epsilon_{3,t} \; . \tag{2.5}$$

In summary we find strong evidence of value (Israel et al., 2018), short-term equity momentum spill-over (Gebhardt et al., 2005), bond reversal (Chordia et al., 2017) and quality (Israel et al., 2018; Houweling & van Zundert, 2017). High profile factors like downside risk by Bai et al. (2019), bond long-term reversal by Bali, Subrahmanyam, and Wen (2021), bond momentum by Jostova et al. (2013) and illiquidity by Bao et al. (2011) do not replicate in our sample. For our main results in Table 2.5 we use CRSP/Compustat matched and default treated data. Therefore, bonds that do not have equity/financial statement information in CRSP/Compustat and bond observations after a positive default flag are removed. For further clarity, we report the results from bi-variate sorts from alternative datasets in Appendix 2.A. In Tables 2.11 and 2.12, we report results using (1) not CRSP/Compustat matched and (2) not CRSP/Compustat matched and not default treated data, respectively. In Table 2.13, we report data for CRSP/Compustat matched, but not default treated data. The not CRSP/Compustat matched dataset includes a higher proportion of high yield and less liquid bonds while the not default treated dataset includes also defaulted bonds. Results are in general stable across the different datasets, however the different characteristics of the datasets raise some interesting points which we discuss below.

Starting from *bond characteristics*, we see that none of the Age, Life, Mdur and Offer-amt appear to be significant. The sign of Mdur is negative as reported in the literature by Israel et al. (2018), Kelly, Palhares, and Pruitt (2020), however, is not significant after controlling for rating. In our alternative datasets, we see that the impact of Mdur is stronger among high yield and when we do not exclude defaults. We think that higher risk-adjusted returns for lower duration bonds are more likely to be compensation for refinancing risk rather than indication of a quality factor as indicated by Israel et al. (2018). Size proxied by Offer-amt also appears to be positive but statistically insignificant similar to Houweling and van Zundert (2017).

Downside risk variables appear to be insignificant despite the recent findings of Bai et al. (2019). Looking at results from the alternative datasets, we find that VaR and ES are affected by the treatment of defaults. In our dataset, we remove bonds after they default, however Bai et al. (2019) do not directly control for defaulted bonds. Kelly et al. (2020) using data from Interactive Data Corporation (ICE) data (ex Merrill Lynch Fixed Income Indexes) also does not detect the VaR factor. We think these results offer support to our argument since its common in the methodology of benchmark indexes to include only performing (non-defaulted) bonds.

For *intangibles* we find that a factor formed by going long/short the companies with the highest/lowest operating accruals (OA) produces a positive premium. The sign is opposite to the one found in equities, however, we think the sign is in line with a risk reward paradigm for bonds. Accruals are a measure of cash flow quality. Companies with low cash flow quality carry a higher credit risk and should offer higher returns to compensate investors. Chordia et al. (2017) similarly report a positive sign but insignificant results. For the *investment* category we find no significant results across our variables. These results are in contrast to a negative relationship of investment and expected corporate bond returns reported by Chordia et al. (2017) and Choi and Kim (2018). For *profitability* variables, we report positive and significant signs for $\Delta(\Delta\mathrm{GM} - \Delta\mathrm{S})$, while EPS, ROA and ROE are significant only after controlling for MKT and TERM effects. These findings point to the direction of a quality factor as reported by Israel et al. (2018), Houweling and van Zundert (2017) and Kelly et al. (2020).

For *past returns* we find strong evidence for equity momentum spillover and bond reversal similar to the literature. However, we fail to replicate the classic momentum introduced by Jostova et al. (2013) using bond's total returns. We are aware that our sample is dominated by investment grade bonds and bond momentum is reported to exist only in high yield, however, when adjusting our sample to include only high yield, we still do not find any significant relationship. Our results are in line with Kelly et al. (2020) that also fail to replicate momentum effects in corporate bonds. Finally, we find reversal effects for spread, i.e. going long/short the bonds that have experienced the largest spread increase/decrease produces a positive premium.

For *trade friction* variables we find no significant evidence. The biggest surprise is the illiquidity factor reported by Bao et al. (2011), where despite getting the correct positive sign, results appear to be not significant after controlling for rating. Finally for variables in the *value* category, we get strong positive and significant results for the value factor, Value^{ips}, suggested by Israel et al. (2018). The value factor suggested by Houweling and van Zundert (2017), Value^{hvz}, has the correct sign but is not statistically significant. Other value variables like C2D, E2P and Sales$_g$ appear to be positive and significant after controlling for MKT and TERM.

We are aware that performing statistical testing for over sixty different characteristics creates a multiple testing bias. The multiple testing bias has been well known for many years in the statistics literature, however, its impact on false discoveries in finance has been recently highlighted by Harvey et al. (2016). Standard ways for controlling for multiple testing is the Bonferroni correction and the Benjamini and Hochberg (1995) (BH). The Bonferonni correction controls for the family-wise-error-rate (FWER) while the BH controls for false-discovery-rate (FDR). However, applying those methods is not straightforward as the correlation of the multiple hypothesis test significantly impacts results. For an extreme example consider that if the hypothesis tests were perfectly correlated, no multiple testing adjustment would be necessary. For practical reasons we apply the two $|t|$-cutoffs of 2.78 and 3.39 proposed by Harvey et al. (2016) and applied by Hou et al. (2020) in the most recent and comprehensive equity anomaly replication study. Using the lower of the two proposed cutoffs, we see that r^e_{1-0}, r^b_{1-0}, $\Delta(\Delta\text{GM} - \Delta\text{S})$ and Value^{ips} are significant. Using the stricter 3.39 limit, only r^b_{1-0} survives. However, looking at the different datasets, all 4 factors mentioned above appear to be rather stable, which gives us further indication that they are true discoveries.

Despite the usefulness and ease of applicability of portfolio sorts, there are also clear limitations. Using portfolio sorts, it quickly becomes cumbersome to control for more than one or two variables. Assuming we live in a multivariate SDF world, this puts significant limitations. A straightforward way to extend portfolio sorts – to accommodate for a larger set of variables – is to use predictive regressions. As mentioned by Cochrane (2011) and demonstrated by Freyberger et al. (2020), “Portfolio sorts are really the same thing as non-parametric cross-sectional regressions, using non-overlapping histogram weights.”

Table 2.5: **Factors from bi-variate sorts (CRSP/Compustat matched).** For each characteristic we first create quantile portfolios according to rating and then according to the characteristic of interest. Factors are constructed as the difference between the high and low quantile portfolios across all rating portfolios. In the first column we report the mean and respective t-stat, whereas in the rest of the columns we report the intercept from regressions of each factor on MKT, MKT + TERM and MKT + TERM + DEF, respectively. Newey and West (1987) HAC robust standard errors, lag = 3 months. We denote 5% and 1% significance level with * and **, respectively. Data from October-2005 to December-2018.

	Factor	α_1	α_2	α_3
β^b	0.17	-0.10	-0.03	-0.03
	0.51	-0.96	-0.33	-0.32
Bond Characteristics:				
Age	0.03	-0.01	-0.01	-0.01
	0.35	-0.11	-0.19	-0.20
Life	0.00	-0.21	-0.01	-0.01
	0.00	-2.08	-0.10	-0.09
Mdur	-0.09	-0.29**	-0.09	-0.09
	-0.42	-2.80	-1.35	-1.38
Offer-amt	0.02	0.02	0.07	0.07
	0.16	0.27	0.99	1.20
Downside risk:				
ES10	0.25	-0.03	0.05	0.05
	0.65	-0.25	0.41	0.40
VaR10	0.17	-0.10	0.00	0.00
	0.48	-0.73	0.02	0.02
VaR5	0.22	-0.05	0.02	0.02
	0.63	-0.41	0.16	0.16
Intangibles:				
AOA	-0.04	-0.03	0.03	0.03
	-0.58	-0.45	0.50	0.55
OA	0.10*	0.10*	0.12**	0.12**
	2.55	2.40	2.78	2.78
OL	-0.09	-0.07	0.01	0.01
	-0.95	-0.65	0.17	0.17
TAN	0.13	0.04	-0.01	-0.01
	0.81	0.41	-0.15	-0.16
Investment:				
ΔCEQ	-0.07	0.02	0.12	0.12
	-0.52	0.27	1.39	1.60
ΔPI2A	-0.05	0.00	0.03	0.03
	-0.42	0.01	0.42	0.44
ΔShout	0.17	0.10	0.01	0.01
	1.37	1.19	0.14	0.14
INV	0.03	0.09	0.12	0.12
	0.25	1.05	1.35	1.35
IVC	0.01	0.02	0.04	0.04
	0.22	0.48	0.75	0.75
Returns:				
Sprd[6]	0.54*	0.38*	0.34*	0.34*
	2.05	2.27	2.10	2.16
r^e_{12-2}	0.02	0.12	0.16	0.16
	0.11	1.05	1.44	1.46
r^b_{12-2}	-0.12	0.02	0.12	0.12
	-0.41	0.10	0.69	0.70

	Factor	α_1	α_2	α_3
r^e_{12-7}	-0.01	0.04	0.07	0.07
	-0.09	0.38	0.83	0.82
r^b_{12-7}	-0.14	-0.05	-0.03	-0.03
	-0.54	-0.28	-0.21	-0.21
r^e_{1-0}	0.63**	0.61**	0.59**	0.59**
	3.04	3.33	3.09	3.15
r^b_{1-0}	-0.40**	-0.39**	-0.35**	-0.35**
	-3.31	-3.14	-2.83	-2.85
r^e_{2-1}	0.08	0.10	0.17	0.17
	0.77	1.01	1.83	1.81
r^b_{2-1}	0.19	0.23	0.19	0.19
	1.15	1.57	1.50	1.52
r^e_{36-13}	-0.14	-0.10	-0.08	-0.08
	-1.32	-1.02	-0.78	-0.77
r^b_{36-13}	-0.19	-0.18	-0.15	-0.15
	-1.45	-1.34	-1.26	-1.27
r^e_{6-2}	0.07	0.15	0.16	0.16
	0.49	1.43	1.61	1.63
r^b_{6-2}	0.03	0.13	0.22	0.22
	0.14	0.80	1.41	1.45
Profitability:				
ATO	-0.01	0.04	0.06	0.06
	-0.07	0.50	0.70	0.70
CTO	-0.15	-0.08	0.03	0.03
	-1.34	-0.86	0.38	0.41
$\Delta(\Delta\text{GM} - \Delta\text{S})$	0.22**	0.24**	0.25**	0.25**
	2.58	3.29	3.78	4.09
EPS	0.11	0.19	0.22*	0.22*
	0.72	1.72	2.05	2.07
IPM	0.07	0.14	0.15	0.15
	0.45	1.26	1.43	1.43
NOA	-0.03	0.04	0.05	0.05
	-0.21	0.45	0.67	0.70
PCM	0.13	0.18	0.18	0.18
	0.91	1.44	1.72	1.73
PM	0.10	0.17	0.16	0.16
	0.60	1.22	1.42	1.43
Prof	0.03	0.07	0.07	0.07
	0.25	0.82	0.86	0.90
RNA	-0.01	0.07	0.14	0.14
	-0.08	0.77	1.54	1.53
ROA	0.10	0.20	0.22*	0.22*
	0.57	1.92	2.31	2.32
ROC	0.05	0.09	0.10	0.10
	0.53	1.39	1.40	1.39

	Factor	α_1	α_2	α_3
ROE	0.08	0.17	0.20*	0.20*
	0.46	1.56	1.99	1.98
ROIC	-0.11	-0.01	0.01	0.01
	-0.66	-0.11	0.07	0.07
S2C	-0.14	-0.06	0.04	0.04
	-1.22	-0.72	0.41	0.50
SAT	-0.12	-0.07	0.02	0.02
	-1.18	-0.66	0.24	0.25
Trade Frictions:				
AT	0.10	0.03	-0.02	-0.02
	1.09	0.46	-0.19	-0.20
Trade^b_{vol}	0.08	0.07	0.07	0.07
	1.21	1.25	1.24	1.11
Illiq	0.25	0.11	0.13	0.12
	1.09	0.98	1.15	1.21
LME	-0.01	0.02	-0.01	-0.01
	-0.05	0.19	-0.06	-0.06
L_{turn}	0.15	0.09	0.07	0.07
	1.01	0.86	0.69	0.69
Value:				
A2ME	0.16	0.00	-0.04	-0.04
	0.61	0.02	-0.32	-0.33
BEME	-0.01	-0.06	0.00	0.00
	-0.04	-0.58	0.04	0.06

	Factor	α_1	α_2	α_3
C	0.15	0.08	0.03	0.03
	1.11	0.90	0.29	0.34
C2D	0.04	0.14	0.17*	0.17*
	0.29	1.61	2.11	2.12
D2P	0.15	0.01	-0.07	-0.07
	0.63	0.04	-0.49	-0.50
ΔSO	0.18	0.10	0.03	0.03
	1.40	1.34	0.36	0.38
E2P	0.18	0.24*	0.28*	0.28*
	1.25	1.96	2.16	2.17
NOP	-0.11	-0.08	0.00	0.00
	-1.33	-0.99	0.03	0.04
O2P	-0.04	-0.02	0.01	0.01
	-0.62	-0.36	0.15	0.14
Q	0.07	0.12	0.11	0.11
	0.85	1.67	1.60	1.67
S2P	0.10	0.01	0.01	0.01
	0.52	0.10	0.12	0.13
Sales_g	0.08	0.17	0.29**	0.29**
	0.68	1.89	2.96	3.03
Value^{hvz}	0.31	0.17	0.14	0.14
	1.59	1.82	1.46	1.52
Value^{ips}	0.66**	0.53**	0.46**	0.46**
	2.95	3.97	3.92	4.20

2.5 Choosing factors

2.5.1 The (adaptive) Lasso

In our paper, we are interested in choosing factors that are able to predict expected excess returns. The conditional expectation of excess returns $E(R_{i,t+1}|C_{i,t})$ has the following linear interpretation in the form of the following panel regression:

$$R_{i,t+1} = x'_{i,t}\beta + \epsilon_{i,t+1}, \tag{2.6}$$

where $R_{i,t+1}$ is bond i's excess return in time period $t+1$, $x'_{i,t} = [1\ C_{i,t}]$ and $C_{i,t}$ equals the vector of p bond, equity and financial characteristics. The error is denoted by $\epsilon_{i,t+1}$. If we denote $y = R_{i,t+1}$ and $\mathbf{X} = x'_{i,t}$ we can write the OLS estimator of equation (2.6) compactly using the following equation:

$$\hat{\beta}_{ols} = \arg\min_{\beta} \left(\frac{1}{N} ||y - \mathbf{X}\beta||_2^2 \right), \tag{2.7}$$

where N refers to the number of total bond-month observations, $||y-\mathbf{X}\beta||_2^2 = \sum_{\forall t}\sum_{\forall i}(R_{i,t+1} - x'_{i,t}\beta)^2$. The Lasso estimator is then defined by Tibshirani (1996) as:

$$\hat{\beta}_{lasso}(\lambda) = \arg\min_{\beta} \left(\frac{1}{N} ||y - \mathbf{X}\beta||_2^2 + \lambda ||\beta||_1 \right), \tag{2.8}$$

where $\lambda \geq 0$ is a penalty parameter and $||\beta||_1 = \sum_{j=1}^{p} |\beta_j|$. The tuning parameter λ is typically determined by cross-validation. Zou (2006) further introduce the adaptive Lasso, which has a different penalization term, that allows the weights to vary for each parameter. This is achieved by having the assigned individual weights being inversely proportional to a first stage β estimate. The adaptive Lasso has the same advantage that Lasso has: It can shrink some of the coefficients to exactly zero, performing thus a selection of variables due to the regularization. Lasso methods are therefore well suited to our particular research question. The adaptive Lasso of Zou (2006) can be expressed with the following formula:

$$\hat{\beta}_{ada-lasso}(\lambda) = \arg\min_{\beta} \left(\frac{1}{N} ||y - \mathbf{X}\beta||_2^2 + \lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\hat{\beta}_{init,j}|} \right), \tag{2.9}$$

where the parameters $\hat{\beta}_{init}$ are determined in a first stage. There are several different ways how $\hat{\beta}_{init}$ is set. E.g. Zou (2006) suggest the use of the OLS estimator, $\hat{\beta}_{init} = \hat{\beta}_{ols}$. If multicollinearity is an issue, Ridge regression can also be considered as a viable alternative to the OLS. Bühlmann and van de Geer (2011) set $\hat{\beta}_{init} = \hat{\beta}_{lasso}$.

In our analysis, we follow Messmer and Audrino (2017) in adopting an adaptive Lasso which has been shown to be superior in terms of selecting true characteristics from a large set of potential characteristics. This is in part due to certain scenarios where the ("regular") Lasso is inconsistent for variable selection. The adaptive Lasso on the other hand, enjoys the oracle properties. Namely, it performs as well as if the true underlying model were given in advance (Zou, 2006). To determine the value of the tuning parameter λ, we employ a ten-fold cross-validation scheme. We estimate the parameters $\hat{\beta}_{init}$ in the first stage with a Ridge regression. For both the Ridge regression and the subsequent adaptive Lasso estimation we use the "one standard error" rule, i.e. we choose the most parsimonious model whose error is no more than one standard error above the error of the best model.

2.5.2 Data preparation

Before implementing the Lasso we have to deal with two issues. Firstly, we need to correct for extreme and often erroneous observations in the set of characteristics. To do so, we winsorize independent variables at the 5% level cross-sectionally. Secondly, to apply the pooled estimation of Lasso, we have to standardize our data in a way that preserves cross-sectional information but adjusts for time variation. Similar to portfolio sorts, our analysis focuses explicitly on the cross-sectional dimension of expected returns, essentially disregarding any time series aspects. As such, we standardize equity and bond characteristics cross-sectionally to have mean 0 and standard deviation of 1. To better understand the last issue raised, consider any characteristic that combines financial data with market data such as book-to-market, earnings-to-price, or debt-to-market capitalization. Those characteristics will fluctuate over time partly due to firm-specific reasons but also due to market-wide movements. So we would expect to see lower book-to-market and debt-to-market capitalization across the cross-section when equity markets are strong. Standardizing at each point in time neutralizes the time dimension while preserving the cross-sectional ranking. For the dependent variable, i.e. the excess returns, we cross-sectionally winsorize at the 5% level and then demean. Even though our excess returns do not have substantial outliers, we still find that the stability of the Lasso results is improved, i.e. selection of variables remains the same even due to the randomness arising from the cross-validation procedure in order to set the tuning parameters.

2.5.3 Empirical results

In Table 2.6 we report the coefficients of the chosen model according to the adaptive Lasso. The chosen model is a ten factor model with Value^{ips}, r^b_{1-0} and r^e_{1-0} being the top three factors in terms of the size of the coefficient. The remaining factors of the chosen model are Mdur, r^b_{6-2}, Sprd^6, VaR5, r^b_{36-13}, A2ME and Rating. Contrary to equity studies, financial variables from Compustat do not appear to have strong power in predicting corporate bond returns. The adaptive Lasso results broadly align with the portfolio sort analysis, finding value, bond reversal, and equity momentum spill-over as the dominant factors. The adaptive Lasso proposes a model of larger dimension than the one proposed by the portfolio sorts. However, we note that portfolio sorts are inferior in detecting factors since we only control for a single variable.

In terms of the signs of the coefficients, the results appear in line with initial results from portfolio sorts and existing literature. We find a positive relationship for value and equity momentum spillover as reported by Israel et al. (2018) and Gebhardt et al. (2005) and a negative relationship for the bond reversal factor as first reported by Chordia et al. (2017). The Mdur has a negative sign indicating that shorter duration bonds earn higher returns than longer duration bonds. Momentum, spread reversal and VaR5 all have a positive sign as reported by Jostova et al. (2013), Israel et al. (2018) and Bai et al. (2019), respectively. Finally, bond long term reversal (r^b_{36-13}) has a negative sign as in Bali et al. (2021) and rating has a positive sign meaning that lower-rated companies earn higher returns, which is in line with a risk-reward paradigm and the credit risk factor in Bai et al. (2019).

Table 2.6: **Characteristics chosen by the adaptive Lasso**. The table exhibits the regression results for the adaptive Lasso, sorted by the absolute value of the coefficients. Coefficient values are scaled by a factor of 1000. Data from October-2005 to December-2018.

Characteristic	Coefficient
r^e_{1-0}	1.47
Value^{ips}	0.90
r^b_{1-0}	-0.56
Mdur	-0.49
r^b_{6-2}	0.44
Sprd^6	0.44
VaR5	0.26
r^b_{36-13}	-0.18
A2ME	-0.15
Rating	0.07

2.5.4 Representative bond

Allowing for multiple bonds for a single firm can lead to biases because we implicitly focus on firms with more outstanding bonds. In addition, the number of outstanding bonds typically correlates to the firm's overall size. As a robustness check, we estimate the adaptive Lasso using the representative bond approach, i.e., we select a single bond for each firm each month. We follow Haesen, Houweling, and van Zundert (2017) and Israel et al. (2018) to define the criteria applied to choose the representative bond, a detailed description of the filtering criteria can be seen in Table 2.7. In essence, the selection procedure aims to identify a sample of liquid and cross-sectionally comparable bonds for each firm. Steps 1 to 4 in Table 2.7 select bonds on the basis of (1) seniority, (2) maturity, (3) age and (4) size. Step 5 counters practical problems that arise if bonds are of the same issue size. A direct consequence of using the representative bond approach is that it reduces the sample size. Following the steps outlined in Table 2.7 to select a representative bond, our sample size reduces from originally 7,839 issues of 1,087 firms with a total of 227,795 bond-month observations to 1,650 issues of 1,087 firms with a total of 23,765 bond-month observations.

Table 2.7: **Selection procedure for a representative bond**. The table displays the procedure to select a representative bond for each issuer, each month. Bonds are selected on the basis of 1) seniority, 2) maturity, 3) age and 4) size with the goal of creating a liquid and cross-sectionally comparable sample.

Step	Decision rule
1	For each issuer only keep those bonds in which the rating corresponds to the largest fraction of debt outstanding
2	If issuer has bonds with time-to-maturity between 5 and 15 years, remove all other bonds for that issuer from the sample
3	If issuer has bonds that are at most 2 years old, remove all other bonds for that issuer
4	Finally, pick the bond with the largest amount outstanding
5	(Bonds with same amounts outstanding will still remain; in this case drop the one with shorter time-to-maturity)

Once we construct the representative bond sample, we then apply the predictive regression framework and the adaptive Lasso methodology described above. In Table 2.8 we report the characteristics and coefficients of the chosen characteristics. As we see our initial results using the full sample of bonds remain robust. The characteristics chosen are the same as the main analysis of our paper in Table 2.6. The only difference that we observe is that despite the fact that modified-duration is chosen, the coefficient is significantly smaller in magnitude. We think this is a reasonable outcome since we restrict bonds to have time to maturity between 5 and 15 years. In that sense we are reducing the variability of the modified-duration characteristic which leads to a weaker coefficient.

Table 2.8: **Characteristics and their coefficients chosen by the adaptive Lasso (representative bond).** The table exhibits the regression results for the adaptive Lasso, sorted by the absolute value of the coefficients. Coefficient values are scaled by a factor of 1000. Data from October-2005 to December-2018.

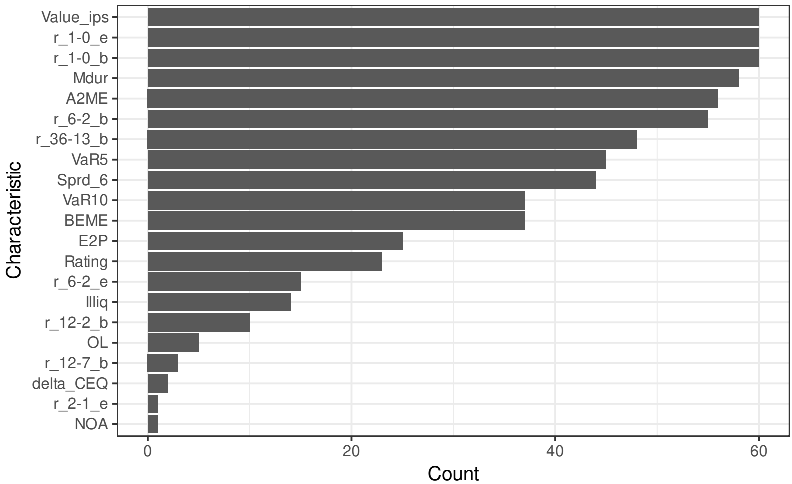
Characteristic	Coefficient
r^e_{1-0}	1.55
Value^{ips}	0.74
r^b_{1-0}	-0.63
Sprd^6	0.47
r^b_{6-2}	0.42
VaR5	0.28
r^b_{36-13}	-0.23
A2ME	-0.11
Rating	0.10
Mdur	-0.06

2.5.5 Out-of-sample performance

We test our adaptive Lasso approach on an out-of-sample expanding window setup in order to provide a practical application of our results and observe the variability of the chosen factors. Due to the comparatively large number of observations needed to perform machine learning exercises, we use 50 percent of the sample observations as the initial training sample. Specifically, 50 percent of the sample data is located prior to December 31st, 2013. Therefore, data up until December 2013 serve as our initial training sample, where we estimate the adaptive Lasso and build predictions for excess returns starting in January 2014. Predictions for January 2014 are then formed based on the values of the independent variables from December 2013. The initial training sample is appended month by month until November 2018. In total, this gives us 60 out-of-sample forecasts.

Figure 2.2 displays the frequency of the characteristics that were chosen over the 60 out-of-sample forecasts, which can also be interpreted as a robustness check. To this end, we can examine how the selected characteristics vary as the sample composition changes. Ideally, the ten characteristics from our full model in Table 2.6 are also frequently chosen in building our out-of-sample forecasts. We can see that this is mostly the case, with the top nine chosen characteristics in the out-of-sample exercise corresponding to the ones selected in our full sample model. Most importantly, the top three characteristics of our full sample adaptive Lasso, r^e_{1-0}, Value^{ips} and r^b_{1-0}, are chosen in each of

Figure 2.2: **Frequency of selected characteristics**. In this figure we display the frequency that each characteristic is choosen during the 60 out-of-sample forecasts. Data from January-2014 to December-2018.



the 60 out-of-sample forecasts. We think that these results confirm the validity of our full sample adaptive Lasso and the importance of r^e_{1-0}, Value^{ips} and r^b_{1-0} for the cross-section of corporate bonds.

To evaluate the economic benefit for an agent who invests according to the predictions of the adaptive Lasso, we build a composite factor (F^{lasso}) by going long/short the bonds with forecasted returns in the top/bottom 20%. To evaluate the performance of the F^{lasso}, we perform two types of analysis. First, we test whether F^{lasso} can be explained by existing models. We use six competing models and run corresponding intercept tests. In our analysis, we include the four-factor model proposed by Bai et al. (2019) and the four-factor model proposed by Houweling and van Zundert (2017). For DRF, CRF, LRF, and REV proposed by Bai et al. (2019) we use our sample and construct the factors from bi-variate sorted portfolios. For the size (SIZE), low-risk (LR), value (VALUE), and momentum (MOM) of Houweling and van Zundert (2017) we use the publicly available data from the authors.[7]

[7] We note that the published factors of Houweling and van Zundert (2017) are long-only and created separately for investment grade and high yield: https://www.robeco.com/en/insights/2018/12/data-sets-factor-investing-in-corporate-bonds.html.

The baseline models that we consider are the following: (1) MKT, (2) MKT, TERM, (3) MKT, TERM, DEF, (4) MKT, TERM, DRF, CRF, LRF, REV, (5) MKT, TERM, SIZE^{ig}, LR^{ig}, VALUE^{ig}, MOM^{ig}, and (6) MKT, TERM, SIZE^{hy}, LR^{hy}, VALUE^{hy}, MOM^{hy}.

Table 2.9 displays the results of the intercept tests. The adaptive Lasso factor generates significant monthly excess returns of 0.28%. The factor survives all six intercept tests. The risk-adjusted excess return of the factor, expressed by the intercepts, is significant and relatively stable, ranging from 0.24% to 0.32% even though we test for a large set of competing models. Our results confirm that F^{lasso} carries a positive and significant premium which can not be explained by existing models.

Table 2.9: **Intercept test**. In this table panel we report the intercept and t-stat from the regression of F^{lasso} on each of the six baseline models. Newey and West (1987) HAC robust standard errors, lag = 3 months. We denote 5% and 1% significance level with * and **.

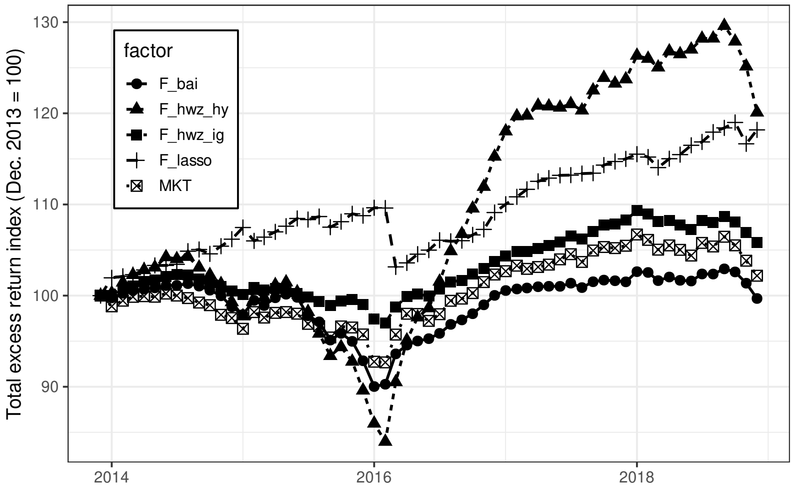
	Factor	α_1	α_2	α_3	α_4	α_5	α_6
F^{lasso}	0.28*	0.30*	0.31**	0.31**	0.24*	0.23*	0.32**
	2.37	2.47	2.98	2.92	2.02	2.51	3.66

We further study the risk/return characteristics of the F^{lasso} and the potential benefits for an investor that follows this strategy. We compare the F^{lasso} against MKT and the corresponding multi-factor portfolios from Bai et al. (2019) and Houweling and van Zundert (2017), which we construct by forming equal-weight portfolios of the respective factors.[8] In Table 2.10a we report annualized mean return, volatility, Sharpe ratio, skewness, kurtosis and beta to the market for the different strategies. As we see, the F^{lasso} achieves significantly higher returns than MKT, F^{bai} and F$^{hwz}_{ig}$ and marginally lower returns than F$^{hwz}_{hy}$. However, we need to keep in mind that our sample is dominated by investment grade bonds while F$^{hwz}_{hy}$ is a long only factor constructed only from high yield rated bonds, so it is expected to have higher returns. In Figure 2.3 we show the cumulative performance of the different factors, confirming the above observations. Looking at Sharpe ratios, we see that F^{lasso} achieves almost seven times the Sharpe ratio of the market and double the Sharpe ratio of F$^{hwz}_{ig}$ and F$^{hwz}_{hy}$.

We continue our analysis by performing Sharpe ratio tests between F^{lasso} and

[8]For example F^{bai} is the equal-weight portfolio of DRF, CRF, REV and LRF.

Figure 2.3: **Cumulative performance of competing factors**. In this figure we display the cumulative performance of the five competing factors, namely F^{lasso}, MKT, F^{bai} and F^{hwz}_{ig} and F^{hwz}_{hy}. We construct multi-factor portfolios by forming equal-weight portfolios of the respective factors. Data from January-2014 to December-2018.



the respective baseline strategy. Given the limited size of our sample and the non-normal characteristics of the factors, we employ the bootstrap method of Ledoit and Wolf (2008) to perform inference, see Appendix 2.B. The null hypothesis is that the F^{lasso} and the respective baseline strategy have the same Sharpe ratio. We perform hypothesis testing by constructing a bootstrap interval at a 95% confidence level. It follows that the test rejects the null hypothesis if zero is not contained in the interval. In Table 2.10b, we report the difference of monthly Sharpe ratios and the corresponding upper and lower bounds. For reference we also report the upper/lower bound from the delta-method. As expected, the bootstrapped interval is more conservative. Despite the considerable improvement achieved from F^{lasso}, we see that the results are not significant with the lower bound of the Sharpe ratio difference being negative across both methods.

We think that the discrepancy between the intercept and Sharpe ratio test is due to the short length of our data set and the increased volatility/outliers of the particular period we examine. As we see from Table 2.10a, all the competing strategies have kurtosis well above 3, which indicate the presence of heavy tails. The kurtosis of F^{lasso} at 21.84 is much higher than the rest of the factors and its due to a single observation. In this case, we argue that the intercept test is more informative and has higher precision because part of the volatility is explained from the factor loadings. We explore this hypothesis by simply removing the month observation with the biggest draw-down across all strategies. In that case, we find that F^{lasso} leads to significant Sharpe ratio improvement. We thus conclude that the diverging results are indeed due to outliers.

Table 2.10: **Performance statistics**. In the left panel we report annualized performance statistics. For the competing factor models of Bai et al. (2019) and Houweling and van Zundert (2017) we create single factors by equal weighting each factor. In the right panel we report monthly Sharpe ratio differentials and upper/lower confidence interval bounds (UB*/LB*) at 95% level according to the bootstrap method of Ledoit and Wolf (2008). Block length=5. We denote the delta-method upper/lower bounds as UB/LB. Data from January-2014 to December-2018.

	F^{lasso}	MKT	F^{bai}	F^{hwz}_{ig}	F^{hwz}_{hy}
Mean	3.41	0.49	-0.00	1.15	3.92
Vol	3.62	3.58	3.48	2.12	7.22
SR	0.94	0.14	-0.00	0.54	0.54
Skew	-3.65	0.21	-0.07	-0.05	0.57
Kurt	21.84	4.56	6.06	3.26	4.74
β^{mkt}	-0.33	-	0.86	0.55	1.72
t-stat	-1.30	-	11.60	19.92	12.68

(a) Annualized performance statistics

	F^{lasso}	MKT	F^{Bai}	F^{hwz}_{ig}	F^{hwz}_{hy}
ΔSR	-	0.23	0.27	0.12	0.12
LB*	-	-0.60	-0.71	-0.65	-0.72
UB*	-	1.06	1.25	0.88	0.95
LB	-	-0.24	-0.23	-0.34	-0.39
UB	-	0.71	0.78	0.57	0.62

(b) Sharpe ratio

2.6 Conclusion

In this paper, we study the cross-section of corporate bonds. We employ a large dataset of bond, equity and financial characteristics and utilize the adaptive Lasso in order to choose the most relevant characteristics for the cross-section of corporate bonds. In the first part of our analysis, we use portfolio sorts to get a first idea of our data and replicate famous factors documented in the literature. Interestingly, we find that downside risk factors proposed by Bai et al. (2019), the momentum factor Jostova et al. (2013) and illiquidity factor Bao et al. (2011) fail to replicate in our sample. Our results about the downside risk factors are in line with the findings of Kelly et al. (2020). Our hypothesis is that downside risk factors are sensitive to the treatment of defaulted bonds. The results of the alternative datasets in Appendix 2.A support such a hypothesis. From the portfolio sorts analysis the dominant factors are the value factor proposed by Israel et al. (2018) and the one-month bond reversal and equity momentum spill-over proposed by Chordia et al. (2017) and Gebhardt et al. (2005), respectively. We also find strong evidence for a quality factor proxied by $\Delta(\Delta\text{GM} - \Delta\text{S})$.

While portfolio sorts are easy to implement and have appealing attributes, they quickly suffer from the curse of dimensionality. For that purpose we utilize the adaptive Lasso in a predictive regression setting. When running the adaptive Lasso on our full dataset, the chosen model is a ten factor model with Value^{ips}, r^b_{1-0} and r^e_{1-0} being the top three factors in terms of size of the coefficient. The remaining factors of the chosen model are Mdur, r^b_{6-2}, Sprd^6, VaR5, r^b_{36-13}, A2ME and Rating. The adaptive Lasso results are broadly in line with the portfolio sort analysis, finding value, bond reversal and equity momentum spill-over as the dominant factors. The adaptive Lasso proposes a model of larger dimension than the one proposed by portfolio sorts, however, we note that portfolio sorts are inferior, since they control only for a single variable at a time.

In order to validate our results and test the usefulness of adaptive Lasso, we run an out-of-sample exercise, where we use an expanding window approach and forecast the returns of the whole cross-section each month. The out-of-sample results validate the selected characteristics from our full sample analysis. The top nine chosen characteristics in the out-of-sample exercise correspond to those selected in our full sample model. Most importantly, the top three characteristics of our full sample adaptive Lasso, r^e_{1-0}, Value^{ips} and r^b_{1-0}, are chosen in each of the 60 out-of-sample forecasts. We then construct a factor portfolio by going long/short the bonds with forecasted returns in the top/bottom 20%. The adaptive Lasso factor generates significant monthly excess returns of 0.28%. Using intercept tests we show that F^{lasso} carries a positive and significant premium which can not be explained by existing models. The risk-adjusted excess return of the factor, expressed by the intercepts, is significant and relatively stable, ranging from 0.24% to 0.32% even though we test for a large set of competing models. We thus conclude that the adaptive Lasso framework can deliver superior returns for an investor who invests according to the forecasted returns of the model.

References

Abarbanell, J. S., & Bushee, B. J. (1997). Fundamental Analysis, Future Earnings, and Stock Prices. *Journal of Accounting Research*, *35*(1), 1–24.

Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, *59*(3), 817–858.

Bai, J., Bali, T. G., & Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, *131*(3), 619–642.

Balakrishnan, K., Bartov, E., & Faurel, L. (2010). Post loss/profit announcement drift. *Journal of Accounting and Economics*, *50*(1), 20–41.

Bali, T. G., Subrahmanyam, A., & Wen, Q. (2021). Long-term reversals in the corporate bond market. *Journal of Financial Economics*, *139*(2), 656–677.

Ball, R., Gerakos, J., Linnainmaa, J. T., & Nikolaev, V. V. (2015). Deflating profitability. *Journal of Financial Economics*, *117*(2), 225–248.

Bandyopadhyay, S. P., Huang, A. G., & Wirjanto, T. S. (2010). *The Accrual Volatility Anomaly* (Working Paper). University of Waterloo.

Bao, J., Pan, J., & Wang, J. (2011). The Illiquidity of Corporate Bonds. *The Journal of Finance*, *66*(3), 911–946.

Basu, S. (1977). Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. *The Journal of Finance*, *32*(3), 663–682.

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of Financial Economics*, *12*(1), 129–156.

Bektić, D., Wenzler, J.-S., Wegener, M., Schiereck, D., & Spielmann, T. (2019). Extending Fama–French Factors to Corporate Bond Markets. *The Journal of Portfolio Management*, *45*(3), 141–158.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bessembinder, H., Kahle, K. M., Maxwell, W. F., & Xu, D. (2009). Measuring Abnormal Bond Performance. *The Review of Financial Studies*, *22*(10), 4219–4258.

Bhandari, L. C. (1988). Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence. *The Journal of Finance*, *43*(2), 507–528.

Boudoukh, J., Michaely, R., Richardson, M., & Roberts, M. R. (2007). On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing. *The Journal of Finance*, *62*(2), 877–915.

Brown, D. P., & Rowe, B. (2007). *The Productivity Premium in Equity Returns* (SSRN Working Paper No. 993467).

Bühlmann, P., & van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Berlin, Heidelberg: Springer.

Chandrashekar, S., & Rao, R. K. S. (2009). *The Productivity of Cash and the Cross-Section of Expected Stock Returns* (McCombs Research Paper Series No. FIN-03-09).

Choi, J., & Kim, Y. (2018). Anomalies and market (dis)integration. *Journal of Monetary Economics*, *100*, 16–34.

Chordia, T., Goyal, A., Nozawa, Y., Subrahmanyam, A., & Tong, Q. (2017). Are Capital Market Anomalies Common to Equity and Corporate Bond Markets? An Empirical Investigation. *Journal of Financial and Quantitative Analysis*, *52*(4), 1301–1342.

Cochrane, J. H. (2011). Presidential Address: Discount Rates. *The Journal of Finance*, *66*(4), 1047–1108.

Cooper, M. J., Gulen, H., & Schill, M. J. (2008). Asset Growth and the Cross-Section of Stock Returns. *The Journal of Finance*, *63*(4), 1609–1651.

Correia, M., Richardson, S., & Tuna, İ. (2012). Value investing in credit markets. *Review of Accounting Studies*, *17*(3), 572–609.

D'Acunto, F., Liu, R., Pflueger, C., & Weber, M. (2018). Flexible prices and leverage. *Journal of Financial Economics*, *129*(1), 46–68.

Datar, V. T., Y. Naik, N., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, *1*(2), 203–219.

Davis, J. L., Fama, E. F., & French, K. R. (2000). Characteristics, Covariances, and Average Returns: 1929 to 1997. *The Journal of Finance*, *55*(1), 389–406.

Dick-Nielsen, J. (2009). Liquidity Biases in TRACE. *The Journal of Fixed Income*, *19*(2), 43–55.

Dick-Nielsen, J. (2014). *How to Clean Enhanced TRACE Data* (SSRN Working Paper No. 2337908).

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (2008). Dissecting Anomalies. *The Journal of Finance*, *63*(4), 1653–1678.

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance*, *75*(3), 1327–1370.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting Characteristics Nonparametrically. *The Review of Financial Studies*, *33*(5), 2326–2377.

Gandhi, P., & Lustig, H. (2015). Size Anomalies in U.S. Bank Stock Returns. *The Journal of Finance*, *70*(2), 733–768.

Gebhardt, W. R., Hvidkjaer, S., & Swaminathan, B. (2005). Stock and bond market interaction: Does momentum spill over? *Journal of Financial Economics*, *75*(3), 651–690.

Gorodnichenko, Y., & Weber, M. (2016). Are Sticky Prices Costly? Evidence from the Stock Market. *The American Economic Review*, *106*(1), 165–199.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Haesen, D., Houweling, P., & van Zundert, J. (2017). Momentum spillover from stocks to corporate bonds. *Journal of Banking & Finance*, *79*, 28–41.

Hahn, J., & Lee, H. (2009). Financial Constraints, Debt Capacity, and the Cross-section of Stock Returns. *The Journal of Finance*, *64*(2), 891–921.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the Cross-Section of Expected Returns. *The Review of Financial Studies*, *29*(1), 5–68.

Haugen, R. A., & Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, *41*(3), 401–439.

Hirshleifer, D., Kewei Hou, Teoh, S. H., & Yinglei Zhang. (2004). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, *38*, 297–331.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating Anomalies. *The Review of Financial Studies*, *33*(5), 2019–2133.

Houweling, P., & van Zundert, J. (2017). Factor Investing in the Corporate Bond Market. *Financial Analysts Journal*, *73*(2), 100–115.

Israel, R., Palhares, D., & Richardson, S. (2018). Common factors in corporate bond returns. , *16*(2), 17–46.

Jostova, G., Nikolova, S., Philipov, A., & Stahel, C. W. (2013). Momentum in Corporate Bond Returns. *The Review of Financial Studies*, *26*(7), 1649–1693.

Kelly, B. T., Palhares, D., & Pruitt, S. (2020). *Modeling Corporate Bond Returns* (SSRN Working Paper No. 3720789).

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian Investment, Extrapolation, and Risk. *The Journal of Finance*, *49*(5), 1541–1578.

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859.

Lewellen, J. (2015). The Cross-section of Expected Stock Returns. *Critical Finance Review*, *4*(1), 1–44.

Litzenberger, R. H., & Ramaswamy, K. (1979). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics*, *7*(2), 163–195.

Lyandres, E., Sun, L., & Zhang, L. (2008). The New Issues Puzzle: Testing the Investment-Based Explanation. *The Review of Financial Studies*, *21*(6), 2825–2855.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, *29*(2), 449–470.

Messmer, M., & Audrino, F. (2017). *The (adaptive) Lasso in the Zoo - Firm Characteristic Selection in the Cross-Section of Expected Returns* (SSRN Working Paper No. 2930436).

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708.

Novy-Marx, R. (2011). Operating Leverage. *Review of Finance*, *15*(1), 103–134.

Ou, J. A., & Penman, S. H. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, *11*(4), 295–329.

Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics*, *104*(1), 162–185.

Politis, D. N., & Romano, J. P. (1992). A Circular Block Resampling Procedure for Stationary Data. In R. LePage & L. Billard (Eds.), *Exploring the Limits of Bootstrap* (pp. 263–270). New York: John Wiley.

Pontiff, J., & Woodgate, A. (2008). Share Issuance and Cross-sectional Returns. *The Journal of Finance*, *63*(2), 921–945.

Richardson, S. A., Sloan, R. G., Soliman, M. T., & Tuna, İ. (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics*, *39*(3), 437–485.

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, *11*(3), 9–16.

Sloan, R. G. (1996). Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings? *The Accounting Review*, *71*(3), 289–315.

Soliman, M. T. (2008). The Use of DuPont Analysis by Market Participants. *The Accounting Review*, *83*(3), 823–853.

Thomas, J. K., & Zhang, H. (2002). Inventory Changes and Future Returns. *Review of Accounting Studies*, *7*(2), 163–187.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429.

2.A Alternative datasets

Table 2.11: **Factors from bi-variate sorts (Not CRSP/Compustat matched)**. For each characteristic we first create quantile portfolios according to rating and then according to the characteristic of interest. Factors are constructed as the difference between the high and low quantile portfolios across all rating portfolios. In the first column we report the mean and respective t-stat, whereas in the rest of the columns we report the intercept from regressions of each factor on MKT, MKT+TERM and MKT+TERM+DEF, respectively. Newey and West (1987) HAC robust standard errors, lag = 3 months. We denote 5% and 1% significance level with * and **, respectively. Data from October-2005 to December-2018.

	Factor	α_1	α_2	α_3
β^b	0.21	-0.10	-0.09	-0.14
	0.61	-0.56	-0.39	-0.58
Bond Characteristics:				
Age	0.20*	0.14	0.08	0.06
	1.99	1.72	0.83	0.61
Life	0.12	-0.13	-0.03	-0.04
	0.42	-1.09	-0.19	-0.33
Mdur	-0.12	-0.34**	-0.20*	-0.20*
	-0.50	-3.07	-2.00	-2.22
Offer-amt	-0.09	-0.04	0.04	0.09
	-0.62	-0.33	0.33	0.74
Downside risk:				
ES10	0.34	0.03	0.02	-0.02
	0.93	0.19	0.11	-0.12
VaR10	0.30	-0.00	-0.01	-0.06
	0.83	-0.01	-0.08	-0.34
VaR5	0.36	0.05	0.01	-0.04
	0.98	0.35	0.08	-0.23
Returns:				
Sprd^6	0.72*	0.53*	0.42	0.39
	2.15	1.99	1.64	1.55
r^b_{12-2}	-0.03	0.11	0.26	0.35
	-0.10	0.47	1.18	1.65

	Factor	α_1	α_2	α_3
r^b_{12-7}	0.14	0.20	0.26	0.31
	0.78	1.20	1.71	1.99
r^b_{1-0}	-0.82*	-0.76*	-0.54	-0.52
	-2.16	-2.05	-1.79	-1.75
r^b_{2-1}	0.37	0.37	0.11	0.06
	1.12	1.25	0.44	0.22
r^b_{36-13}	-0.18	-0.19	-0.24	-0.23
	-1.34	-1.50	-1.94	-1.79
r^b_{6-2}	0.05	0.18	0.35	0.44
	0.23	0.82	1.86	2.34
Trade Frictions:				
Trade^b_{vol}	-0.01	-0.01	-0.01	0.04
	-0.14	-0.20	-0.23	0.52
Illiq	0.40	0.22	0.14	0.09
	1.53	1.42	1.18	0.81
Value:				
Value^{hvz}	0.35	0.20	0.12	0.07
	1.58	1.66	0.94	0.59
Value^{ips}	0.75**	0.63**	0.58**	0.56**
	2.92	3.09	2.82	2.61

Table 2.12: **Factors from bi-variate sorts (Not CRSP/Compustat matched/Not Altman matched).** For each characteristic we first create quantile portfolios according to rating and then according to the characteristic of interest. Factors are constructed as the difference between the high and low quantile portfolios across all rating portfolios. In the first column we report the mean and respective t-stat, whereas in the rest of the columns we report the intercept from regressions of each factor on MKT, MKT+TERM and MKT+TERM+DEF, respectively. Newey and West (1987) HAC robust standard errors, lag = 3 months. We denote 5% and 1% significance level with * and **, respectively. Data from October-2005 to December-2018.

	Factor	α_1	α_2	α_3
β^b	0.21	-0.23	-0.20	-0.17
	0.65	-1.36	-1.01	-0.91
Bond Characteristics:				
Age	0.21*	0.10	0.05	0.07
	2.28	1.52	0.52	0.97
Life	0.10	-0.26*	-0.15	-0.13
	0.36	-2.40	-1.28	-1.24
Mdur	-0.27	-0.55**	-0.36**	-0.36**
	-1.05	-3.77	-2.94	-2.90
Offer-amt	-0.02	0.06	0.16	0.10
	-0.13	0.53	1.40	1.56
Downside risk:				
ES10	0.53	0.09	0.10	0.13
	1.42	0.61	0.66	0.85
VaR10	0.48	0.04	0.00	0.04
	1.32	0.27	0.01	0.24
VaR5	0.51	0.07	0.03	0.07
	1.40	0.51	0.21	0.49
Returns:				
Sprd^6	0.80*	0.52	0.36	0.39
	2.18	1.84	1.43	1.60
r^b_{12-2}	-0.21	-0.02	0.14	0.10
	-0.77	-0.10	0.67	0.49
r^b_{12-7}	0.01	0.10	0.21	0.19
	0.03	0.65	1.41	1.31
r^b_{1-0}	-0.91*	-0.77*	-0.47	-0.54*
	-2.19	-2.09	-1.56	-1.95
r^b_{2-1}	0.34	0.30	0.07	0.14
	1.17	1.23	0.27	0.79
r^b_{36-13}	-0.23	-0.25*	-0.33**	-0.31**
	-1.51	-2.07	-2.71	-2.67
r^b_{6-2}	0.02	0.18	0.33	0.29
	0.10	0.86	1.74	1.58
Trade Frictions:				
Trade^b_{vol}	0.10	0.11	0.11	0.07
	1.48	1.61	1.73	1.45
Illiq	0.49	0.23	0.16	0.19
	1.78	1.73	1.45	1.89
Value:				
Value^{hvz}	0.57*	0.33**	0.27*	0.29*
	2.21	2.73	2.20	2.54
Value^{ips}	0.78**	0.58**	0.54**	0.55**
	2.79	3.10	2.81	3.01

Table 2.13: **Factors from bi-variate sorts (CRSP/Compustat matched/Not Altman matched).** For each characteristic we first create quantile portfolios according to rating and then according to the characteristic of interest. Factors are constructed as the difference between the high and low quantile portfolios across all rating portfolios. In the first column we report the mean and respective t-stat, whereas in the rest of the columns we report the intercept from regressions of each factor on MKT, MKT+TERM and MKT+TERM+DEF, respectively. Newey and West (1987) HAC robust standard errors, lag = 3 months. We denote 5% and 1% significance level with * and **, respectively. Data from October-2005 to December-2018.

	Factor	α_1	α_2	α_3
β^b	0.25	-0.05	-0.00	-0.00
	0.77	-0.51	-0.01	-0.04
Bond Characteristics:				
Age	0.10	0.06	0.05	0.04
	1.13	0.81	0.79	0.55
Life	-0.02	-0.27	-0.07	-0.08
	-0.08	-2.93	-1.05	-1.19
Mdur	-0.15	-0.38**	-0.18**	-0.19**
	-0.69	-3.97	-2.74	-2.78
Offer-amt	0.06	0.06	0.09	0.13
	0.68	0.81	1.59	1.86
Downside risk:				
ES10	0.37	0.04	0.11	0.08
	0.96	0.34	0.89	0.76
VaR10	0.28	-0.02	0.06	0.04
	0.81	-0.20	0.56	0.38
VaR5	0.34	0.03	0.08	0.05
	0.96	0.28	0.71	0.51
Intangibles:				
AOA	0.00	-0.00	0.07	0.05
	0.01	-0.05	1.13	0.85
OA	0.10*	0.10*	0.12**	0.12**
	2.27	2.25	2.80	2.82
OL	-0.11	-0.09	-0.01	-0.04
	-1.10	-0.90	-0.10	-0.63
TAN	0.26	0.16	0.10	0.10
	1.58	1.59	0.99	0.95
Investment:				
ΔCEQ	-0.07	0.02	0.14	0.14
	-0.52	0.18	1.51	1.50
ΔPI2A	-0.11	-0.06	-0.03	-0.03
	-1.00	-0.81	-0.41	-0.39
ΔShout	0.25	0.16	0.08	0.07
	1.83	1.88	0.91	0.83
INV	-0.09	-0.00	0.07	0.07
	-0.74	-0.04	0.84	0.90
IVC	-0.03	-0.01	0.01	0.01
	-0.53	-0.20	0.27	0.13
Returns:				
Sprd[6]	0.58*	0.39*	0.33	0.32
	2.11	2.18	1.89	1.84
r^e_{12-2}	-0.06	0.06	0.10	0.12
	-0.31	0.51	0.98	1.13
r^b_{12-2}	-0.14	0.02	0.14	0.21
	-0.44	0.08	0.75	1.33

	Factor	α_1	α_2	α_3
r^e_{12-7}	-0.06	-0.01	0.04	0.04
	-0.51	-0.11	0.44	0.49
r^b_{12-7}	-0.17	-0.08	-0.05	0.01
	-0.69	-0.45	-0.34	0.08
r^e_{1-0}	0.64**	0.62**	0.59**	0.55**
	3.09	3.41	3.16	3.25
r^b_{1-0}	-0.41**	-0.40**	-0.35**	-0.34**
	-3.54	-3.33	-2.94	-2.83
r^e_{2-1}	0.04	0.07	0.15	0.12
	0.46	0.63	1.62	1.33
r^b_{2-1}	0.23	0.27	0.24	0.24
	1.29	1.78	1.83	1.76
r^e_{36-13}	-0.19	-0.14	-0.10	-0.12
	-1.66	-1.32	-0.92	-1.09
r^b_{36-13}	-0.26*	-0.25	-0.21	-0.19
	-1.97	-1.86	-1.79	-1.42
r^e_{6-2}	0.05	0.14	0.16	0.17
	0.36	1.32	1.51	1.61
r^b_{6-2}	0.07	0.17	0.27	0.33*
	0.31	0.99	1.69	2.29
Profitability:				
ATO	-0.05	0.00	0.02	0.01
	-0.53	-0.00	0.25	0.11
CTO	-0.20	-0.14	-0.02	-0.04
	-1.92	-1.44	-0.21	-0.49
$\Delta(\Delta\text{GM} - \Delta\text{S})$	0.22**	0.25**	0.25**	0.25**
	3.00	3.74	4.04	4.14
EPS	0.06	0.14	0.19	0.19*
	0.44	1.46	1.88	2.15
IPM	0.02	0.10	0.11	0.12
	0.11	0.93	1.18	1.41
NOA	-0.07	0.00	0.03	0.04
	-0.54	0.03	0.30	0.47
PCM	0.12	0.17	0.18	0.20*
	0.90	1.52	1.90	2.20
PM	0.08	0.15	0.15	0.17
	0.49	1.14	1.42	1.61
Prof	0.01	0.06	0.08	0.07
	0.13	0.76	1.06	0.96
RNA	-0.04	0.05	0.13	0.11
	-0.30	0.56	1.50	1.25
ROA	0.04	0.15	0.19*	0.19*
	0.24	1.58	2.11	2.29
ROC	0.02	0.06	0.08	0.06
	0.24	1.09	1.22	1.11

	Factor	α_1	α_2	α_3
ROE	0.01	0.11	0.17	0.16
	0.08	1.12	1.71	1.82
ROIC	-0.11	0.01	0.02	0.02
	-0.65	0.06	0.26	0.24
S2C	-0.16	-0.08	0.02	0.02
	-1.40	-0.94	0.23	0.22
SAT	-0.13	-0.09	0.01	-0.02
	-1.49	-0.90	0.05	-0.21
Trade Frictions:				
AT	0.10	0.03	-0.02	0.00
	1.06	0.38	-0.17	-0.00
Trade^{b}_{vol}	0.11	0.10	0.09	0.13*
	1.68	1.71	1.65	2.41
Illiq	0.34	0.18	0.18	0.15
	1.51	1.61	1.73	1.58
LME	-0.12	-0.08	-0.08	-0.07
	-0.90	-0.74	-0.89	-0.79
L_{turn}	0.13	0.04	0.01	0.02
	0.80	0.46	0.09	0.19
Value:				
A2ME	0.28	0.10	0.03	0.03
	1.07	0.78	0.27	0.29
BEME	0.01	-0.05	-0.01	0.02
	0.07	-0.58	-0.06	0.24

	Factor	α_1	α_2	α_3
C	0.20	0.12	0.06	0.04
	1.50	1.49	0.77	0.54
C2D	-0.00	0.10	0.15	0.14
	-0.02	1.22	1.78	1.91
D2P	0.25	0.08	-0.02	-0.02
	1.01	0.59	-0.13	-0.19
ΔSO	0.27*	0.18*	0.11	0.11
	2.04	2.38	1.32	1.36
E2P	0.12	0.19	0.24*	0.24*
	0.87	1.65	1.98	2.18
NOP	-0.14	-0.11	-0.03	-0.04
	-1.71	-1.34	-0.26	-0.40
O2P	-0.08	-0.06	-0.02	-0.04
	-1.27	-0.92	-0.29	-0.58
Q	0.05	0.10	0.10	0.08
	0.63	1.52	1.47	1.36
S2P	0.17	0.04	0.06	0.02
	0.76	0.33	0.51	0.25
Sales_g	0.05	0.14	0.27**	0.26**
	0.39	1.47	2.71	2.67
Value^{hvz}	0.47*	0.30**	0.26*	0.22*
	2.17	2.92	2.44	2.07
Value^{ips}	0.61**	0.46**	0.39**	0.37**
	2.68	3.52	3.28	3.10

2.B Bootstrap inference for Sharpe ratio testing

Consider two investment strategies i and j with excess returns R_{it} and R_{jt}. The mean vector μ and the covariance matrix Σ are given by:

$$\mu = \begin{bmatrix} \mu_i \\ \mu_j \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_i^2 & \sigma_{ij} \\ \sigma_{ji} & \sigma_j^2 \end{bmatrix}. \tag{2.10}$$

The difference between the Sharpe ratio of the two strategies is given by:

$$\Delta = \mathrm{Sh}_i - \mathrm{Sh}_j = \frac{\mu_i}{\sigma_i} - \frac{\mu_j}{\sigma_j}. \tag{2.11}$$

Let $\gamma_i = \mathrm{E}[R_{it}^2]$ and $\gamma_j = \mathrm{E}[R_{jt}^2]$ be the uncentered second moments and $\upsilon = (\mu_i, \mu_j, \gamma_i, \gamma_j)$. Then we can express the Sharpe ratio difference Δ as a function of υ ($\Delta = f(\upsilon)$), with $f(a,b,c,d) = (a/\sqrt{c-a^2}) - (b/\sqrt{d-b^2})$. If we assume that $\sqrt{T}(\hat{\upsilon} - \upsilon) \xrightarrow{d} N(0, \Psi)$ then from the the delta method it follows that:

$$\sqrt{T}(\hat{\Delta} - \Delta) \xrightarrow{d} N(0, \nabla' f(\upsilon) \Psi \nabla f(\upsilon)). \tag{2.12}$$

The standard error of $\hat{\Delta}$ can be written as:

$$S(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\upsilon) \hat{\Psi} \nabla f(\upsilon)}{T}}, \tag{2.13}$$

where $\hat{\Psi}$ is a consistent estimator of Ψ and its typically estimated through HAC robust estimation procedure such as Newey and West (1987) or Andrews (1991). However, it is well-known that such HAC inference is often too optimistic when tails are heavier than normal or sample sizes are small. Ledoit and Wolf (2008) propose to test $H_0 : \Delta = 0$ by inverting a symmetric studentized bootstrap confidence interval, with confidence level $1 - \alpha$. It follows that the null is rejected if zero is not contained in this interval. The process they propose is the following:

First we generate bootstrap data by re-sampling with replacement individual pairs of the two returns series. If there is serial correlation we can apply the block bootstrap of Politis and Romano (1992). We then estimate $\hat{\Delta}^*$, which is the Sharpe ratio difference estimated from the bootstrapped data. We repeat this procedure for B times. Let $z^*_{|\cdot|,\lambda}$ be the λ quantile of the $|\hat{\Delta}^* - \hat{\Delta}|/S(\hat{\Delta}^*)$ random variable, then the $1 - \alpha$ bootstrap confidence interval is given by:

$$\hat{\Delta} \pm z^*_{|\cdot|,1-\alpha}\, S(\hat{\Delta}) \tag{2.14}$$

where $\hat{\Delta}$ is estimated from the original data and $S(\hat{\Delta})$ is estimated from the original data according to equation (2.13). When data are heavy tailed and the sample is small, then $z^*_{|\cdot|,1-\alpha}$ is more conservative than its asymptotic counterpart $z_{1-\alpha}$.

2.C Variable definition

β^b: Beta of each bond with the market using last 36 months of returns. A bond needs to have at least 24 observations out of 36 months in order to be included in our calculation of the β^b.

Bond Char.:

Age: Number of months since the issuance of the bond.

Life: Number of months until maturity of the bond.

Mdur: Modified duration of the bond.

Offer-amt: Offering amount of the bond in million USD.

Downside risk:

ES10: Defined by Bai et al. (2019) as the average of the four lowest monthly return observations over the past 36 months (beyond the 10% VaR threshold).

VaR10: Defined by Bai et al. (2019) as the fourth lowest monthly return observation over the past 36 months. We then multiply the original measure by -1 for convenience of interpretation.

VaR5: Defined by Bai et al. (2019) as the second lowest monthly return observation over the past 36 months. We then multiply the original measure by -1 for convenience of interpretation.

Intangibles:

AOA: We follow Bandyopadhyay, Huang, and Wirjanto (2010) and define AOA as absolute value of operation accruals (OA) which we define below.

OA: We follow Sloan (1996) and define operating accruals as changes in non-cash working capital minus depreciation (DP) scaled by lagged total assets (AT). Non-cash working capital is the difference between non-cash current assets and current liabilities (LCT), debt in current liabilities (DLC) and income taxes payable (TXP). Non-cash current assets are current assets (ACT) minus cash and short-term investments (CHE).

OL: Operating leverage is the sum of cost of goods sold (COGS) and selling, general, and administrative expenses (XSGA) over total assets (AT) as in Novy-Marx (2011).

TAN: We follow Hahn and Lee (2009) and defne tangibility as (0.715 $\times$ total receivables (RECT) + 0.547 $\times$ inventories (INVT) + 0.535 $\times$ property, plant and equipment (PPENT) + cash and short-term investments (CHE)) / total assets (AT).

Investment:

ΔCEQ: We follow Richardson, Sloan, Soliman, and Tuna (2005) in the definition of the percentage change in the book value of equity (CEQ).

ΔPI2A: We define the change in property, plants, and equipment following Lyandres, Sun, and Zhang (2008) as changes in property, plants, and equipment (PPEGT) and inventory (INVT) over lagged total assets (AT).

ΔShout: We follow Pontiff and Woodgate (2008) in the definition of the percentage change in shares outstanding (SHROUT).

INV: We define investment as the percentage year-on-year growth rate in total assets (AT) following Cooper, Gulen, and Schill (2008).

IVC: We define IVC as change in inventories (INVT) between $t-2$ and $t-1$ over the average total assets (AT) of years $t-2$ and $t-1$ following Thomas and Zhang (2002).

Returns:

Sprd[6]: Mom. 6m log(Spread) from Israel et al. (2018). Difference in logs of current spread and spread 6 months ago. Spread is defined as the difference of yield-to-maturity of a bond and its corresponding maturity-matched treasury yield.

r^e_{12-2}: Equity momentum. Total return from 12 months up until 2 months ago.

r^b_{12-2}: Bond momentum. Total return from 12 months up until 2 months ago.

r^e_{12-7}: Equity momentum. Total return from 12 months up until 7 months ago.

r^b_{12-7}: Bond momentum. Total return from 12 months up until 7 months ago.

r^e_{1-0}: Equity momentum. Total return over the past month.

r^b_{1-0}: Bond momentum. Total return over the past month.

r^e_{2-1}: Equity momentum. Total return from 2 months up until 1 month ago.

r^b_{2-1}: Bond momentum. Total return from 2 months up until 1 month ago.

r^e_{36-13}: Equity momentum. Total return from 36 months up until 13 months ago.

r^b_{36-13}: Bond momentum. Total return from 36 months up until 13 months ago.

r^e_{6-2}: Equity momentum. Total return from 6 months up until 2 months ago.

r^b_{6-2}: Bond momentum. Total return from 6 months up until 2 months ago.

Profitability:

ATO: Net sales over lagged net operating assets as in Soliman (2008). Net operating assets are the difference between operating assets and operating liabilities. Operating assets are total assets (AT) minus cash and short-term investments (CHE), minus investment and other advances (IVAO). Operating liabilities are total assets (AT), minus debt in current liabilities (DLC), minus long-term debt (DLTT), minus minority interest (MIB), minus preferred stock (PSTK), minus common equity (CEQ).

CTO: We follow Haugen and Baker (1996) and define capital turnover as ratio of net sales (SALE) to lagged total assets (AT).

$\Delta(\Delta\text{GM} - \Delta\text{S})$: We follow Abarbanell and Bushee (1997) in the definition of the difference in the percentage change in gross margin and the percentage change in sales (SALE). We define gross margin as the difference in sales (SALE) and costs of goods sold (COGS).

EPS: We follow Basu (1977) and define earnings per share as the ratio of income before extraordinary items (IB) to shares outstanding (SHROUT) as of December $t-1$.

IPM: We define pre-tax profit margin as ratio of pre-tax income (PI) to sales (SALE).

NOA: Net operating assets are the difference between operating assets minus operating liabilities scaled by lagged total assets as in Hirshleifer, Kewei Hou, Teoh, and Yinglei Zhang (2004). Operating assets are total assets (AT) minus cash and short-term investments (CHE), minus investment and other advances (IVAO). Operating liabilities are total assets (AT), minus debt in current liabilities (DLC), minus long-term debt (DLTT), minus minority interest (MIB), minus preferred stock (PSTK), minus common equity (CEQ).

PCM: The price to cost margin is the difference between net sales (SALE) and costs of goods sold (COGS) divided by net sales (SALE) as in Gorodnichenko and Weber (2016) and D'Acunto, Liu, Pflueger, and Weber (2018).

PM: The profit margin is operating income after depreciation (OIADP) over sales (SALE) as in Soliman (2008).

Prof: We follow Ball, Gerakos, Linnainmaa, and Nikolaev (2015) and define profitability as gross profitability (GP) divided by the book value of equity as defined above.

RNA: The return on net operating assets is the ratio of operating income after depreciation to lagged net operating assets (Soliman, 2008). Net operating assets are the difference between operating assets minus operating liabilities. Operating assets are total assets (AT) minus cash and short-term investments (CHE), minus investment and other advances (IVAO). Operating liabilities are total assets (AT), minus debt in current liabilities (DLC), minus long-term debt (DLTT), minus minority interest (MIB), minus preferred stock (PSTK), minus common equity (CEQ).

ROA: Return on assets is income before extraordinary items (IB) to lagged total assets (AT) following Balakrishnan, Bartov, and Faurel (2010).

ROC: Return on capital is the ratio of market value of equity (ME) plus long-term debt (DLTT) minus total assets to cash and short-term investments (CHE) as in Chandrashekar and Rao (2009).

ROE: Return on equity is income before extraordinary items (IB) to lagged book value of equity as in Haugen and Baker (1996).

ROIC: Return on invested capital is the ratio of earnings before interest and taxes (EBIT) less nonoperating income (NOPI) to the sum of common equity (CEQ), total liabilities (LT), and cash and short-term investments (CHE) as in Brown and Rowe (2007).

S2C: Sales to cash is the ratio of net sales (SALE) to cash and short-term investments (CHE) following Ou and Penman (1989).

SAT: We follow Soliman (2008) and define asset turnover as the ratio of sales (SALE) to total assets (AT).

Trade Frictions:

AT: Total assets (AT) as in Gandhi and Lustig (2015).

Trade^b_{vol}: Cumulative trading volume of the bond in a given month.

Illiq: Defined by Bao et al. (2011) as: $\text{ILLIQ}_t = -\text{Cov}_t(\Delta p_{itd}, \Delta p_{itd+1})$, where $\Delta p_{itd} = p_{itd} - p_{itd-1}$ is the log price change of bond i and day d of month t.

LME: Size is the total market capitalization of the previous month defined as price (PRC) times shares outstanding (SHROUT) as in Fama and French (1992).

L_{turn}: Turnover is last month's volume (VOL) over shares outstanding (SHROUT) as in Datar, Y. Naik, and Radcliffe (1998).

Value:

A2ME: We follow Bhandari (1988) and define assets to market capitalization as total assets (AT) over market capitalization as of December $t-1$. Market capitalization is the product of shares outstanding (SHROUT) and price (PRC).

BEME: Ratio of book value of equity to market value of equity. Book equity is shareholder equity (SH) plus deferred taxes and investment tax credit (TXDITC), minus preferred stock (PS). SH is shareholders' equity (SEQ). If missing, SH is the sum of common equity (CEQ) and preferred stock (PS). If missing, SH is the difference between total assets (AT) and total liabilities (LT). Depending on availability, we use the redemption (item PSTKRV), liquidating (item PSTKL), or par value (item PSTK) for PS. The market value of equity is as of December $t-1$. The market value of equity is the product of shares outstanding (SHROUT) and price (PRC). See Rosenberg, Reid, and Lanstein (1985) and Davis, Fama, and French (2000).

C: Ratio of cash and short-term investments (CHE) to total assets (AT) as in Palazzo (2012).

C2D: Cash flow to price is the ratio of income and extraordinary items (IB) and depreciation and amortization (DP) to total liabilities (LT).

D2P: Debt to price is the ratio of long-term debt (DLTT) and debt in current liabilities (DLC) to the market capitalization as of December $t-1$ as in Litzenberger and Ramaswamy (1979). Market capitalization is the product of shares outstanding (SHROUT) and price (PRC).

ΔSO: Log change in the split adjusted shares outstanding as in Fama and French (2008). Split adjusted shares outstanding are the product of Compustat shares outstanding (CSHO) and the adjustment factor (AJEX).

E2P: We follow Basu (1983) and define earnings to price as the ratio of income before extraordinary items (IB) to the market capitalization as of December $t-1$. Market capitalization is the product of shares outstanding (SHROUT) and price (PRC).

NOP: Net payout ratio is common dividends (DVC) plus purchase of common and preferred stock (PRSTKC) minus the sale of common and preferred stock (SSTK) over the market capitalization as of December as in Boudoukh, Michaely, Richardson, and Roberts (2007).

O2P: Payout ratio is common dividends (DVC) plus purchase of common and preferred stock (PRSTKC) minus the change in value of the net number of preferred stocks outstanding (PSTKRV) over the market capitalization as of December as in Boudoukh et al. (2007).

Q: Tobin's Q is total assets (AT), the market value of equity (SHROUT times PRC) minus cash and short-term investments (CEQ), minus deferred taxes (TXDB) scaled by total assets (AT).

S2P: Sales to price is the ratio of net sales (SALE) to the market capitalization as of December following Lewellen (2015).

Sales_g: Sales growth is the percentage growth rate in annual sales (SALE) following Lakonishok, Shleifer, and Vishny (1994).

Value^{hvz}: Defined by Houweling and van Zundert (2017) as calculating cross-sectionally a regression and generate fitted values: $\hat{\text{Spread}}_t = c + \beta^{RAT}\text{RAT}_t + \beta^{MAT}\text{MAT}_t + \beta^{\Delta S}\Delta\text{Spread}_{t,t-3}$. The percentage difference between the actual and the fitted spread is then the value characteristic.

Value^{ips}: Defined by Israel et al. (2018) as calculating cross-sectionally a regression and generate fitted values: $\hat{\text{Spread}}_t = c + \beta^{RAT}\text{RAT}_t + \beta^{DUR}\text{DUR}_t + \beta^{VOL}\text{VOL}_{t,t-12}$. $\text{VOL}_{t,t-12}$ refers to the bond's volatility over the past year. The percentage difference between the actual and the fitted spread is then the value characteristic.

Chapter 3

Luck versus skill in the corporate bond fund market

Co-author: Dimitrios Nteventzis

Abstract

In this paper, we evaluate the ability of U.S. corporate bond fund managers to generate alpha. We apply the False Discovery Rate (FDR) to distinguish between "skill" and "luck." We find that long-term outperformance remains elusive, with only 1% of the funds able to generate significant alpha over their life. However, fund managers are able to generate alpha over the short-term with the proportion of skilled funds increasing to 13.5% when we examine three-year sub-periods. To confirm these findings, we design an out-of-sample investment strategy where we invest in funds according to their estimated "skill" from past returns. Our strategy generates positive and significant alpha, which confirms the persistence in outperformance over the short-run. Our results are economically meaningful for investors suggesting that dynamic and active manager selection pays off.

3.1 Introduction

Are active managers able to outperform? This question is one of the oldest in the field of finance and has profound welfare implications for investors. While there have been numerous studies on the performance of U.S. equity fund managers, there have been far less on U.S. bond funds and only a few for U.S. corporate bond funds. This paper aims to fill the gap in the existing literature by explicitly studying the performance of U.S. corporate bond funds. Most studies so far on the U.S. corporate space focus on the performance of the average fund. Both theoretical and empirical results hint at the fact that funds generate on average zero/negative alpha gross/net of fees (see Berk and Green (2004)). It follows that active management is a zero-sum game, where returns of unskilled managers (negative alpha) and skilled ones (positive alpha) are balanced out. We expand the existing literature by analyzing the performance of individual funds. To this end, we are interested in testing whether skilled managers that consistently outperform the market exist and whether investors can identify them and receive superior returns.

To answer this question, we utilize established methods from the equity fund literature. More specifically, we use the False Discovery Rate (FDR) which was introduced by Storey (2002) and applied by Barras, Scaillet, and Wermers (2010) to evaluate U.S. equity funds. The central issue in studying the performance of individual funds is the multiple testing bias that arises. For example, consider using a statistical significance level of 5%, and counting the number of funds with significant alphas to determine how many managers are skilled. In that case, we should expect that 5% of zero-alpha funds (funds with zero alpha in population) will erroneously appear to possess skill just out of luck. The FDR method is a practical and statistically sound approach to control for false discoveries directly. Despite the sound theoretical framework of the FDR, there has been recently a critique from Andrikogiannopoulou and Papakonstantinou (2019) challenging the applicability of FDR on fund evaluation. Their main argument is that FDR produces biased results that typically underestimate the proportion of non-zero alpha funds if we use realistic noise-to-signal ratios and sample size. We take into account the recent findings on the shortcomings of the FDR in fund evaluation by performing a detailed simulation study and also incorporating the comments of Barras, Scaillet, and Wermers (2019) in their reply on the validity of FDR.

Apart from the statistical methodology used to evaluate fund returns, the choice of the benchmark model is also crucial. In the equity space, there are asset pricing models that are widely accepted, such as the four-factor model of Carhart (1997) that is being used in most equity mutual fund papers. However, there is no broad-based agreement in the corporate bonds space yet. In addition, the lack of readily available bond factors has led researchers to use augmented models from the equity literature to measure the performance of bond funds. We contribute to the existing literature by utilizing corporate bond-specific asset pricing factors. More specifically, our reference model is the one suggested by Bai, Bali, and Wen (2019) that focuses on the downside, credit, and liquidity risk.

From the existing literature, Ayadi and Kryzanowski (2011) and Brooks, Gould, and Richardson (2020) are the papers that are most closely related to our work. Ayadi and Kryzanowski (2011) examine the Canadian bond fund market and use the bootstrap to examine the best/worst funds. They find that bad luck is responsible for the left tail funds but find no real skill for the right tail funds. Brooks et al. (2020) examine the U.S. bond market and independently examine different segments of the bond fund market. They examine the performance of individual funds; however, they do not employ a formal statistical method but rather draw inference by imposing a normal distribution with zero mean on the distribution of individual alphas. To our knowledge, our paper is the first one that focuses on the performance of individual U.S. corporate bond funds and uses the FDR methodology to distinguish skill from luck.

First, we evaluate the performance of funds over the long-run by applying the FDR approach on the full sample (158 funds, 198 months). We find that 1.15%, 12.27%, and 86.58% of the funds generate positive, negative, and zero-alpha. Thus, we conclude that alpha generation is scarce over the long-run, while most of the funds generate zero alpha. Our results are in line with the mutual fund literature. Brooks et al. (2020) examine the U.S. bond mutual fund and find no evidence of true skill. Barras et al. (2010) report that only 0.60% of equity fund managers can generate alpha, while they find that the majority of the funds are zero-alpha.

Second, we evaluate the performance of funds over the short-run. According to the theoretical work of Berk and Green (2004), funds might be able to outperform over the short-run before investors reduce their competitive advantage by increasing inflows. To this end, we continue our analysis focusing on 3-year sub-intervals and treating each fund-period observation as a separate fund. We find that performance is stronger than the full sample analysis, indicating that fund managers can generate alpha over the short-run. We estimate the proportions of positive, negative, and zero-alpha to be 13.52%, 10.25%, 76.23%. Barras et al. (2010) find similar evidence examining the U.S. equity mutual funds and Huij and Derwall (2008) find performance persistence in U.S. bond funds using a dataset from 1990 to 2003.

Third, motivated by the fact that corporate bond funds exhibit outperformance over the short-run, we examine whether investors can use past performance to identify talented managers and create a profitable strategy. We follow Barras et al. (2010) and design a fund selection strategy based on the FDR methodology. We show that we can successfully identify managers that possess skill over the short-run. Our strategy generates an economically and statistically meaningful annualized alpha of 1.75%. We also report the results of a simple strategy that invests every year in the funds with alpha above/below a certain percentile threshold, as in Huij and Derwall (2008). The results are similar to our FDR selection strategy and confirm the hypothesis of short-term performance persistence and skill. However, similar to Barras et al. (2010) we observe that the FDR method cannot meaningfully outperform the naive strategy of portfolio formation according to past alpha. We think that the concentration of skilled funds at the extreme right tail and the increasing alphas the more one moves to the right tail are the two main reasons for the inability of the FDR strategy to outperform. It follows that by simply focusing on an extreme percentile like 90%, one can capture a meaningful proportion of skilled funds. Furthermore, the benefits of identifying skilled funds further on the left side of the distribution are diminishing, given the lower alpha and higher proportion of unskilled funds.

Finally, given the recent critique of Andrikogiannopoulou and Papakonstantinou (2019) on the bias of the FDR methodology, we perform a detailed simulation study to evaluate the performance of FDR in our dataset. We use the median size of our sample T_{med} and calibrate the alpha and residual volatility parameter from our data in line with the response of Barras et al. (2019). While we confirm previous findings of bias in FDR estimates, our simulation results are less pessimistic than Andrikogiannopoulou and Papakonstantinou (2019) and closer to the recent response of Barras et al. (2019). Our simulation results confirm that FDR can be successfully applied in the cross-section of corporate bond mutual funds.

Our paper is organized as follows. The second section provides a detailed literature review of mutual fund performance analysis for equity and bond funds and an overview of the respective methods. In the third section, we present our data, while in sections four and five, we present our empirical results and simulation study.

3.2 Literature review

The majority of existing literature on fund performance has focused on equities. The results are sometimes conflicting depending on the methodology used. However, the prevailing view is that long-term outperformance is scarce. Early papers on mutual fund performance try to identify skill by testing for persistence in fund returns. The idea is that if a manager outperforms due to skill, this should also persist in the future. Carhart (1997) applies a four-factor model including momentum and finds that the momentum factor can explain previous evidence of performance persistence. Kosowski, Timmermann, Wermers, and White (2006) use a bootstrap method to evaluate the performance of individual equity funds. They show that while alpha for the broad industry is negative, in line with the "equilibrium accounting theory," a group of "star" managers can persistently deliver significant alpha. Fama and French (2010) perform a similar study to Kosowski et al. (2006) and conclude that while many managers can generate alpha before costs, only a few can generate alpha after costs. Barras et al. (2010) use the FDR in order to correct for false discoveries. They find that only 0.6% of managers possess skill while 75% of the U.S. equity mutual funds have zero alpha. They also find a significant proportion (14.4%) of skilled investors before 1996 but almost none (0.6%) by 2006. Similar findings of a declining ability to generate alpha have been reported by Pástor, Stambaugh, and Taylor (2015). The dominating belief is that increasing competition and market efficiency is restricting opportunities for outperformance. More recently, Harvey and Liu (2018) proposed a new structural approach to model skilled, unskilled, and zero alpha funds. They find that 10% of the U.S. equity funds generate positive alpha, a finding that is in contrast with the 1% reported by Barras et al. (2010).[1]

While there have been numerous studies on the return of actively managed equity funds, there has been less attention on bond funds and only a few studies on corporate bond fund returns. In one of the first studies focusing on bond funds, Blake, Elton, and Gruber (1993) examine government and corporate

[1] For a detailed literature review on equity mutual fund evaluation, see Cremers, Fulkerson, and Riley (2019).

bond funds jointly. They find negative alphas for the aggregate industry and weak results on performance persistence. Ferson, Henry, and Kisgen (2006) examine government bond funds using the implied stochastic discount factor (SDF) from a continuous-time term structure model and report that the average fund outperforms the benchmark. Huij and Derwall (2008) question the existence of skilled bond fund managers by investigating the persistence of relative performance of funds. Using a larger dataset than previous studies, they show that funds that outperform in the past tend to outperform in the future, offering evidence to support the existence of skilled bond fund managers. Ayadi and Kryzanowski (2011) examine the Canadian bond fund market. In line with previous studies, they find on average negative alphas net of fees. When applying the bootstrap technique of Kosowski et al. (2006) to evaluate individual fund performance, they find no skill for the right tail of alphas. Cici and Gibson (2012) focus on corporate bond funds and decompose active returns in security selection, timing, and style. They report negative results on the abilities of bond fund managers; however, their analysis is limited to the average corporate bond fund. Konstantinov and Fabozzi (2021) examines the European bond mutual fund market across different segments (government, investment grade (IG), high yield (HY), etc.). They find that only funds that invest primarily in government bonds can generate alpha and that alpha generation has declined significantly after the GFC (Global Financial Crisis) in 2008. Brooks et al. (2020) examine the U.S. bond market and independently examine different segments of the bond fund market similar to Konstantinov and Fabozzi (2021). They examine the performance of individual funds; however, they do not employ a formal statistical method but rather draw inference by imposing a normal distribution with zero mean on the distribution of individual alphas. They conclude that the average and individual active manager is not able to generate alpha. The results hold across all bond categories. To our knowledge, this is the first paper that focuses on the performance of individual U.S. corporate bond funds and uses the FDR methodology to distinguish skill from luck.

Answering whether individual managers possess skills has specific methodological challenges since it involves multiple hypothesis testing. The bootstrap technique of Kosowski et al. (2006) and Fama and French (2010) deals with the non-normal distributional characteristics of fund returns but is not able to correct for the multiple hypothesis bias. To limit the bias, the authors restrict their analysis to key points of the cross-section of funds. On the other hand, the FDR method of Barras et al. (2010) allows us to directly control for false discoveries and estimate the proportions of skilled and unskilled managers. Despite the sound theoretical foundations of the FDR method, there are empirical difficulties as highlighted by Andrikogiannopoulou and Papakonstantinou (2019). The authors argue that the low accuracy of estimating fund alphas and the short sample period of individual funds makes the FDR estimation biased, as shown in a detailed simulation study. However, Barras et al. (2019) argue in their response that if we use realistic parameters for alphas and residual volatility and incorporate the additional data available from the original study in 2010, then the bias is significantly smaller than the one reported by Andrikogiannopoulou and Papakonstantinou (2019). Harvey and Liu (2018) in a recent study propose a structural approach that aims to reduce the noise in estimating alphas by pooling information from the cross-sectional distribution of alphas. Christiansen, Grønborg, and Nielsen (2020) perform a detailed empirical study to measure the performance of different methods in fund selection through a detailed simulation study. They adjust the simulation parameters to accommodate for realistically short samples. They find that the FDR and bootstrap method have the most attractive characteristics while the more advanced method of Harvey and Liu (2018) performs worse when applied to small samples. Given that the outperformance of newly proposed methods remains uncertain, we prefer to use the established FDR method and examine its accuracy through a detailed simulation study.

3.3 Mutual fund performance evaluation methodology

3.3.1 Set up

Consider N actively managed mutual funds with excess returns net of fees R_t and K asset pricing factors F_t that form the benchmark model. To evaluate the performance of these funds, we estimate:

$$R_{it} = \alpha_i + \beta_i F_t + \epsilon_{it} \tag{3.1}$$

It follows that a fund manager can generate returns either due to its exposure to systematic risk factors through β_i or due to management skills through α_i. We focus on returns net of fees and define skill as the ability to generate $\alpha_i > 0$ after costs. Accordingly, the N funds can be separated into three categories:

- **Unskilled funds:** funds with managers that are unable to recover trading and management fees through their active management skills ($\alpha_i < 0$).

- **Zero alpha funds:** funds with managers that are able to recover trading and management fees through their active management skills ($\alpha_i = 0$).
- **Skilled funds:** funds with managers that can generate alpha in excess of trading and management fees through their active management skills ($\alpha_i > 0$).

A simplistic approach for estimating the number of skilled/unskilled funds would be to perform N tests with the null hypothesis being $\mathrm{H}_0^i : \alpha_i = 0$. This is the no-luck approach, and the major issue is that it fails to account for false discoveries. Figure 3.1 presents a simplified example in order to demonstrate the empirical implications. In Figure 3.1a we demonstrate the t-statistic distribution of unskilled, zero-alpha and skilled funds. For simplicity, we assume a normal distribution centered across -3, 0 and 3, respectively. In Figure 3.1b we demonstrate the cross-sectional t-statistic distribution, which is a mixture of the three skill group distributions in Figure 3.1a, with the weight of each distribution being equal to the proportion of unskilled, zero alpha and skilled funds in the population ($\pi_A^- = 9\%$, $\pi_0 = 90\%$, $\pi_A^+ = 1\%$). Using a significance level of $\gamma = 5\%$ we would estimate the unskilled funds to be 11% for the total population and the skilled funds to be 3%. The reason for the bias is that we do not account for luck (lucky/unlucky funds with zero-alpha but due to luck appear to have positive/negative alpha).

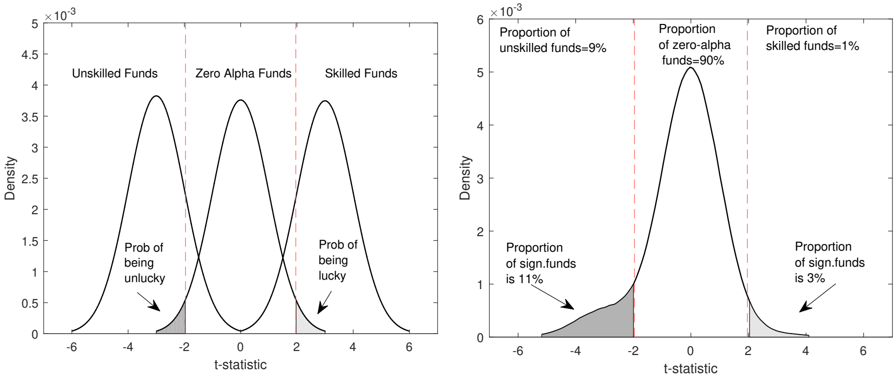
3.3.2 False Discovery Rate (FDR)

The FDR method applied by Barras et al. (2010) offers a robust statistical framework to estimate the proportion of negative, zero, and positive alpha funds while adjusting for false discoveries due to luck. From Figure 3.1a, we see that at a given significance level γ, the probability that a zero-alpha fund is wrongfully accepted as a skilled/unskilled fund is equal to $\gamma/2$. If π_0 is the proportion of zero-alpha funds in population, it follows that the expected proportion of lucky/unlucky funds can be written as:

$$\begin{aligned} E(F_\gamma^+) &= \pi_0 \times \gamma/2 \\ E(F_\gamma^-) &= \pi_0 \times \gamma/2 \ , \end{aligned} \tag{3.2}$$

where $E(F_\gamma^+)$ denotes the expected proportion of lucky funds and $E(F_\gamma^-)$ denotes the expected proportion of unlucky funds. Now to calculate the expected proportion of non-zero funds, we simply need to adjust the observed

Figure 3.1: **Example on bias for multiple testing**. The left panel shows the distribution of the fund t-statistics across the three skill groups (zero alpha, negative and positive). We assume a normal distribution and that the respective distributions are centered around 0 and ± 3. The right panel shows the cross-sectional distribution of the t-statistics, which is a mixture of the three skill group distributions in the left panel, with the weight of each distribution being equal to the proportion of unskilled, zero alpha and skilled funds in the population ($\pi_A^- = 9\%$, $\pi_0 = 90\%$, $\pi_A^+ = 1\%$).



(a) t-statistic distribution by fund category

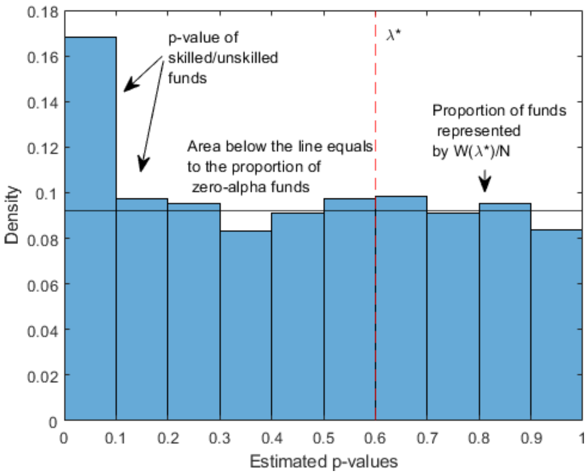
(b) Cross-sectional t-statistic distribution

proportions for the impact of luck. It follows that the proportions of non zero funds can be expressed as:

$$E(T_\gamma^+) = E(S_\gamma^+) - E(F_\gamma^+) = E(S_\gamma^+) - \pi_0 \times \gamma/2$$
$$E(T_\gamma^-) = E(S_\gamma^-) - E(F_\gamma^-) = E(S_\gamma^-) - \pi_0 \times \gamma/2 \ , \tag{3.3}$$

where $E(T_\gamma^+)$ and $E(T_\gamma^-)$ denotes the expected proportions of skilled/unskilled funds and $E(S_\gamma^+)$ and $E(S_\gamma^-)$ denote the proportion of funds with positive/ negative and significant t-statistics in sample.

Figure 3.2: **Distribution of p-values.** The figure displays the p-value histogram of $N = 250$ funds generated from the three skill distributions displayed in Figure 3.1 with probability equal to the proportion of unskilled, zero alpha and skilled funds in the population ($\pi_A^- = 9\%$, $\pi_0 = 90\%$, $\pi_A^+ = 1\%$).



From equation (3.3) we see that the only parameter that we need in order to estimate the true proportions of non-zero funds is the proportion of zero funds, π_0. For that, we can use the estimation approach developed by Storey (2002). Consider we have N hypothesis to test (same as the number of funds) and $\hat{p}_i$ is the p-value of test i. The main two assumptions of the FDR are that: (1) the p-values corresponding to true nulls will be uniformly distributed on [0,1], and (2) the p-values corresponding to the alternative hypothesis will be close to zero. From assumption (2), it follows that there is a threshold $\lambda^* \in (0, 1)$ above which all p-values correspond to true nulls. We employ a bootstrap method to determine λ^* which we describe below. Given that p-values of the null are uniformly distributed, we can estimate the proportion of zero alpha funds ($\hat{\pi}_0$) by counting the number of p-values above λ^*, $W(\lambda^*)$, and extrapolate for the interval $(0, \lambda^*)$. It follows that we can estimate the proportion of zero alpha funds from:

$$\hat{\pi}_0(\lambda^*) = \frac{W(\lambda^*)}{N} \frac{1}{1 - \lambda^*} , \tag{3.4}$$

where $W(\lambda^*)/N$ can be interpreted as the proportion of funds represented by the four rectangles for $\lambda > \lambda^*$. Once we have estimated $\hat{\pi}_0$, we can simply substitute $\hat{\pi}_0$ in equations (3.3) which gives as the proportions of positive and negative alpha after adjusting for false discoveries. Accordingly, the estimated proportions of non zero funds can be written as:

$$\begin{aligned} \hat{\pi}_A^+ &= \hat{S}_\gamma^+ - \hat{F}_\gamma^+ = \hat{S}_\gamma^+ - \hat{\pi}_0 \times \gamma/2 \\ \hat{\pi}_A^- &= \hat{S}_\gamma^- - \hat{F}_\gamma^- = \hat{S}_\gamma^- - \hat{\pi}_0 \times \gamma/2 \end{aligned} \tag{3.5}$$

3.3.3 Bootstrapped p-values

The p-values required to compute $W(\lambda^*)$ in equation (3.4) are estimated using the bootstrap method of Kosowski et al. (2006). Thus, allowing us to take into account the non-normality and asymmetry of funds' returns. To employ the methodology of Kosowski et al. (2006) we first estimate the benchmark model in equation (3.1) and save the estimated parameters $\hat{\beta}_i$ and $\{\hat{\epsilon_{it}}, t = T_{i0}, \ldots, T_{i1}\}$ for each fund. Time indexes $T_{i0}, \ldots, T_{i1}$ denote the first and last observation of fund i. Next, we draw with replacement a pseudo time series of residuals $\{\hat{\epsilon}^b_{it_\epsilon}, t_\epsilon = s^b_{T_{i0}}, \ldots, s^b_{T_{i1}}\}$, where b is an index for the bootstrap simulation run and time indices $s^b_{T_{i0}}, \ldots, s^b_{T_{i1}}$ are drawn uniformly from the original time indices $T_{i0}, \ldots, T_{i1}$. We draw pseudo excess return, R^b_{it}, under the null hypothesis ($\alpha_i = 0$) from:

$$R^b_{it} = \hat{\beta}_i F_t + \hat{\epsilon}^b_{it_\epsilon} \tag{3.6}$$

Then we regress the pseudo returns on the set of original factors:

$$R^b_{it} = \alpha^b_i + \beta^b_i F_t + \tilde{\epsilon}_{it} \tag{3.7}$$

We save α^b_i and the corresponding t-statistic and repeat this procedure for $b = 1, \ldots, 1000$ bootstrap iterations and for all the funds $i = 1, \ldots, N$. Since the distribution of t-statistics might be asymmetric, we follow Davidson and MacKinnon (2004) and estimate the p-value from:

$$\hat{p}_i = 2 \times \min\left(\frac{1}{B}\sum_{b=1}^{B} I\{\widehat{t}^b_i > \widehat{t}_i\}, \frac{1}{B}\sum_{b=1}^{B} I\{\widehat{t}^b_i < \widehat{t}_i\}\right), \tag{3.8}$$

where B is the total number of bootstrap iterations and $I\{\widehat{t}^b_i > \widehat{t}_i\}$ is an indicator function that takes the value one if the bootstrap t-statistic $\widehat{t}^b_i$ is higher than the t-static estimated from the actual sample $\widehat{t}_i$.

3.3.4 Determining the value of λ^* and γ^* from the data

We follow Barras et al. (2010) and determine λ^* and γ^* from the data, utilizing a bootstrap approach. To determine λ^* we first estimate $\hat{\pi}_0(\lambda)$ for a grid of λ values, ($\lambda = 0.30, 0.35, 0.40, .., 0.70$). For each value of λ we form 1,000 bootstrap replications by drawing with replacement from the Nx1 vector of p-values. We estimate the Mean Squared Error (MSE) for each value of λ and choose as λ^* the value that minimizes the MSE, $\lambda^* = \text{argmin}_\lambda \widehat{\text{MSE}}(\lambda)$.

$$\widehat{\text{MSE}}(\lambda) = \frac{1}{1000} \sum_{b=1}^{1000} \left[\hat{\pi}_0^b(\lambda) - \min_\lambda \hat{\pi}_0(\lambda) \right]^2 \tag{3.9}$$

The significance level γ^* is typically determined by the researcher and defines the segment of the tail to be examined for skill/luck. Since our primary focus is to estimate the proportions of funds for the entire population, we need to set a reasonably large value of γ^* in order to keep Type-II error under control. We follow the approach proposed of Barras et al. (2010) to determine γ^* through bootstrap similar to λ^*. First we estimate $\pi_A^-(\gamma)$ from (3.5) using λ^* from the previous step, for a range of γ values ($\gamma = 0.30, 0.35, .., 0.50$). Second, for each value of γ we form 1,000 bootstrap replications by drawing with replacement from the Nx1 vector of p-values. We estimate the Mean Squared Error (MSE) for each value of γ and choose as γ^- the value that minimizes MSE, $\gamma^- = \text{argmin}_\gamma \widehat{\text{MSE}}^-(\gamma)$.

$$\widehat{\text{MSE}}^-(\gamma) = \frac{1}{1000} \sum_{b=1}^{1000} \left[\hat{\pi}_A^{b-}(\gamma) - \max_\gamma \hat{\pi}_A^-(\gamma) \right]^2 \tag{3.10}$$

We use the same approach to choose γ^+. If $\min_\gamma \widehat{\text{MSE}}^-(\gamma) < \min_\gamma \widehat{\text{MSE}}^+(\gamma)$ then we set $\hat{\pi}_A^-(\gamma^*) = \hat{\pi}_A^-(\gamma^-)$ and to preserve the equality, we set $\hat{\pi}_A^+(\gamma^*) = 1 - \hat{\pi}_0 - \hat{\pi}_A^-(\gamma^*)$. Accordingly, if $\min_\gamma \widehat{\text{MSE}}^-(\gamma) > \min_\gamma \widehat{\text{MSE}}^+(\gamma)$, we set $\hat{\pi}_A^+(\gamma^*) = \hat{\pi}_A^+(\gamma^+)$ and $\hat{\pi}_A^-(\gamma^*) = 1 - \hat{\pi}_0 - \hat{\pi}_A^+(\gamma^*)$. We note that in our empirical results we find that simply setting γ^* at a reasonable large value (above 0.20) yields similar results to the bootstrap approach.

3.4 Data description and performance measurement

3.4.1 Mutual fund data

We use the CRSP Survivor Bias-Free Mutual Fund Database for the fund returns. For funds with multiple share classes, we calculate size-weighted returns across all share classes. Excess returns R_{it} are estimated by subtracting the return of the one-month T-Bill. We proceed by filtering our data in order to isolate corporate bond funds. (1) We keep funds with a Lipper classification of A, BBB, GB, HY, IID, SII.[2] (2) We match the fund return series with the available corporate bond factors that span from Nov-2005 to Dec-2018. (3) To avoid incubation bias,[3] we remove funds with history less than 36 months and funds that do not grow more than 50 million USD. (4) We remove Exchange Traded Funds (ETF) using the respective flag provided by the dataset. (5) We remove funds with average corporate allocation below 50%. (6) We remove funds with average equity exposure more than 2.5%. (7) We remove funds with an average allocation of other fixed-income above 30%. (8) We remove short-duration funds based on whether they include the word short in their name. The final dataset includes 198 unique funds and 19,833 observations. In Table 3.2 we describe in detail the filters we apply and the number of observations in each step.

In Table 3.1 we see the descriptive statistics for the full sample and the IG and HY funds separately. The filtering strategy successfully isolates funds that invest primarily in corporate bonds, with the corporate bond allocation being on average above 80%. The average fund in our sample delivers a monthly return of 0.43%, while as expected, returns of HY funds are higher at 0.51% vs. 0.31% for the IG funds. Expense ratios and management fees are also higher for HY funds compared to IG funds. In terms of turnover ratio, we see that the distribution is skewed with some very active funds pushing the mean above 80% while the median of the sample is closer to 60%. Finally in Figure 3.3 we plot the returns of the average IG and HY fund vs. the respective Bloomberg corporate bond index. As we see, both categories are highly correlated to their respective benchmark and generate lower returns over the sample period.

[2] A-Corporate Debt Funds A rated, BBB-Corporate Debt Funds BB rated, GB-General Bond Funds, HY-High Current Yield Funds, IID-Intermediate Intermediate Investment Grade Debt Funds, SII-Short Intermediate Investment Grade Debt Funds.

[3] See Brooks et al. (2020) and Konstantinov and Fabozzi (2021).

Table 3.1: **Descriptive statistics full sample**. In this table we present the descriptive statistics for the full sample of fund returns and characteristics.

	Obs	Funds	Mean	50th	SD	5th	25th	75th	95th
Ret	19,833	198	0.43	0.47	2.01	-2.38	-0.26	1.25	3.08
Size (M USD)	19,833	198	1,459	416	3,004	22	107	1,334	6,466
Exp-ratio %	17,756	178	0.84	0.83	0.34	0.19	0.65	1.03	1.41
Mgmt fee %	17,762	178	0.44	0.48	0.45	0.05	0.37	0.60	0.76
Turn %	17,762	178	92.00	60.00	117.90	21.00	39.00	95.00	263.00
Corp %	13,525	167	84.87	89.75	13.35	56.43	79.33	93.83	97.69

(a) Full sample

	Obs	Funds	Mean	50th	SD	5th	25th	75th	95th
Ret	8,012	88	0.31	0.29	1.36	-1.55	-0.20	0.91	2.18
Size (M USD)	8,012	88	1,358	467	3,095	29	122	1,169	5,789
Exp-ratio %	7,253	81	0.72	0.72	0.34	0.11	0.56	0.96	1.25
Mgmt fee %	7,259	81	0.35	0.40	0.27	0.00	0.30	0.48	0.62
Turn %	7,259	81	105.29	63.00	126.49	18.00	36.00	119.00	325.00
Corp %	5,125	64	79.37	82.35	15.15	50.59	69.08	92.02	97.15

(b) Investment grade

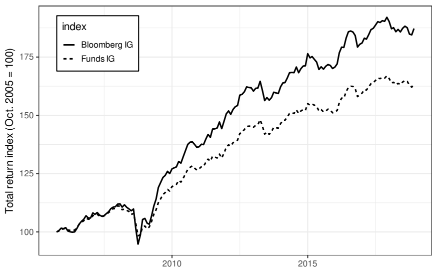
	Obs	Funds	Mean	50th	SD	5th	25th	75th	95th
Ret	11,079	103	0.51	0.67	2.35	-2.73	-0.37	1.49	3.69
Size (M USD)	11,079	103	1,433	354	2,777	19	98	1,447	6,665
Exp-ratio %	9,761	90	0.95	0.89	0.31	0.55	0.76	1.10	1.52
Mgmt fee %	9,761	90	0.50	0.56	0.55	0.16	0.46	0.66	0.80
Turn %	9,761	90	80.17	58.65	109.91	25.00	41.00	82.00	162.00
Corp %	7,893	97	88.91	91.44	9.51	69.64	87.27	94.44	97.97

(c) High yield

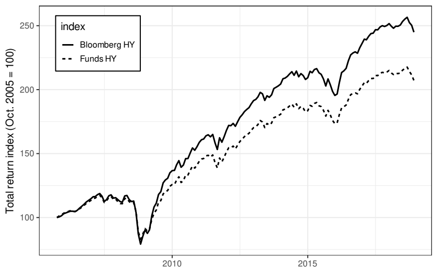
Table 3.2: **Sample set up**. In this table we describe the evolution of our dataset according to the filter/restrictions we apply. Obs refers to the month-fund observation in each sample, while the rest of the variables refer to the number of funds in each respective sample.

Filter	Obs	Num. Funds	IG Funds	HY Funds	IG Obs	HY Obs
Initial sample	189,015	2,866	1,536	623	104,466	41,896
- Keep only IG/HY	146,377	2,127	1,535	623	104,463	41,896
- Match with factor data	120,011	1,917	1365	581	84,367	35,647
- Remove < 36 months	107,664	1,093	789	326	75,987	31,680
- Remove < 50 M USD	100,456	980	706	295	70,560	29,898
- Remove ETF	94,185	894	644	270	65,933	28,254
- Remove corp alloc < 50%	48,918	472	237	249	22,062	26,856
- Remove equity alloc > 2.5%	31,988	323	190	143	17,522	14,466
- Remove other FI alloc > 30%	27,635	279	147	140	13,316	14,319
- Remove short maturity funds	19,833	198	95	110	8,199	11,634

Figure 3.3: **Returns of average IG/HY fund and Bloomberg benchmark index**. In this figure we report the cumulative returns of the average IG/HY fund against the return of the respective benchmark.



(a) IG



(b) HY

3.4.2 Performance measurement and asset pricing factors

This section describes the asset pricing models that we utilize for the performance evaluation of corporate funds. Most of the papers that examine corporate mutual funds' performance have utilized asset pricing models from the equity literature. For example Konstantinov and Fabozzi (2021) and Ayadi and Kryzanowski (2011), both use the Carhart (1997) 4-factor equity model, augmented with a term-spread[4] factor and bond-specific market factor. We

[4]Term-spread is defined as the difference between 30-year U.S. Treasury and 1-month U.S. Treasury.

differentiate from previous papers by utilizing corporate bond-specific asset pricing factors. More specifically, our reference model is the one suggested by Bai et al. (2019). They propose a six-factor model that includes: market (MKT), term-spread (TERM), downside-risk (VAR), liquidity (LIQ), credit risk (CRF) and one month return reversal (REV). We compare our reference model with: (i) the two-factor model of Fama and French (1993) that includes the market factor and term and (ii) the two-factor model augmented with the equity factors[5] of size (SMB), value (HML) and momentum (MOM).

To construct bond factors, we utilize the dataset from TRACE that provides prices and characteristics of the full cross-section of U.S. corporate bonds. Details about the TRACE dataset and cleaning procedure can be found in Appendix 3.B. We construct the bond factors according to Bai et al. (2019). To measure downside-risk, we use the 5% VaR over the last 36 months, which is equivalent to the second-worst return over the same period. Bond illiquidity is measured according to Bao, Pan, and Wang (2011). To be more specific, illiquidity is defined as $\text{ILLIQ}_t = -\text{Cov}_t(\Delta p_{itd}, \Delta p_{itd+1})$, where $\Delta p_{itd} = p_{itd} - p_{itd-1}$ is the log price change of bond i and day d of month t. For the credit risk factor, we use the credit rating and for the one-month reversal factor we use the last month return.

We create traded factors from bi-variate sorts where we use credit rating as the first sorting variable, resulting in 25 (5x5) portfolios. Credit risk is a crucial driver of corporate bond returns; using the rating as a first sorting variable allows us to create portfolios with similar credit risk profiles. To create the credit-risk factor (CRF), we use the VaR 5% as the first sorting variable. We construct each factor as the value-weighted average return difference of the extreme quantile portfolios among all credit rating portfolios. To briefly illustrate how we construct each factor, consider that we have R_{it} excess returns of N test assets, and we want to construct a factor for a single characteristic C_{it}. Then we denote $P_t^{1,5}$ the return at time t of a value-weighted portfolio of assets that belongs in the 1st quantile concerning C_{it} and the 5th quantile concerning credit rating. Then the factor is defined as:

$$\text{Factor}_t = \frac{1}{5}\sum_{j=1}^{5} P_t^{5,j} - \frac{1}{5}\sum_{j=1}^{5} P_t^{1,j} \tag{3.11}$$

To evaluate the performance of the models, we estimate Fama-MacBeth cross-

[5] Data for the equity factors are from the Fama and French website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

sectional regressions for each model. To estimate the beta loading of the factors, we use the full sample available. As we see from results in Table 3.3, only DRF and LIQ are statistically significant from the full model of Bai et al. (2019). Thus, we choose to use a restricted version of the full model that includes market, term, downside risk, and liquidity. We think it is reasonable that CRF and REV are not significant in our sample. First of all, using TRACE data to generate the factors, we find that CRF is not statistically significant in contrast to Bai et al. (2019). Second, while REV is significant in our sample, the reversal phenomenon is attributed to information asymmetry according to Ivashchenko (2019). As such, while one-month reversal appears significant for pricing individual bonds, we think that it is reasonable to assume that long-only funds can not efficiently exploit a market microstructure-driven factor. Moving to the remaining two models that we consider, we see that both perform worse than the 4-factor version of Bai et al. (2019). More specifically, we see that traditional equity factors cannot explain the cross-section of corporate bond funds. These results align with recent literature highlighting the complexity of the cross-section of corporate bonds and the need to develop bond-specific models instead of relying on established models from the equity literature. Given that CRF and REV are not statistically significant in pricing the cross-section of our sample, we continue our analysis using the reduced version of the Bai et al. (2019) model using MKT, TERM, DRF, and LIQ:

$$R_{it} = \alpha_i + \beta_{1,i}\text{MKT}_t + \beta_{2,i}\text{TERM}_t + \beta_{3,i}\text{DRF}_t + \beta_{4,i}\text{LIQ}_t + \epsilon_{3,i,t} \quad (3.12)$$

Table 3.3: **Fama-Macbeth regressions**. In this table we report results from Fama-Macbeth cross-sectional regressions. The standard errors are estimated using the Newey and West (1987) methodology and lag=6. Number of funds=198, T=158. We use *, ** and *** to indicate significance at the 10%, 5%, and 1% levels.

	MKT	TERM	DRF	LIQ	CRF	REV	SMB	HML	MOM	R^2
(1)	0.40**	0.24**	0.64**	0.56*	0.04	0.74	-	-	-	0.61
	2.07	2.05	1.99	1.94	0.18	1.55	-	-	-	
(2)	0.40**	0.12	0.68**	0.45**	-	-	-	-	-	0.62
	2.07	0.96	2.04	1.98	-	-	-		-	
(3)	0.39*	0.11	-	-	-	-	-	-	-	0.54
	1.79	0.89	-	-		-	-		-	
(4)	0.39*	0.15	-	-	-		0.40*	-0.07	-0.34	0.57
	1.81	1.14	-	-	-	-	1.81	-0.22	-0.49	

3.5 Empirical Results

3.5.1 Skill and luck in long-term performance

We begin our analysis by utilizing the entire sample and the FDR methodology to estimate the proportion of skilled/unskilled fund managers. From Table 3.4 we see that only a small percentage of funds can generate alpha, while most of the funds are classified as zero-alpha, meaning they possess just enough skills to retrieve costs but not enough to outperform. We find that 1.15%, 12.27%, and 86.58% of the funds have positive, negative, and zero-alpha, respectively. We conclude that alpha generation is scarce over the long-run, while most of the funds possess some skill to recover trading costs and management expenses. Our results are in line with the mutual fund literature. Brooks et al. (2020) examine the U.S. bond mutual fund and find no evidence of true skill. Barras et al. (2010) report that only 0.60% of equity fund managers can generate alpha, while they also find that the majority of the funds are zero-alpha.

In panel 3.4b of Table 3.4 we demonstrate the impact of luck on identifying skill. In particular, we estimate the proportion of significant alphas resulting from a simple t-statistic comparison, for a grid of significance levels $\gamma = \{0.05, 0.10, 0.20, 0.30\}$. We also report the lucky/unlucky funds and the skilled/unskilled funds resulting from the FDR for the same grid of γ values. To demonstrate the impact let us consider the right tail and the case of $\gamma = 0.20$. If we simply count the funds with significant positive alpha we would erroneously conclude that 10% are able to generate alpha. However in reality 8.66% of those funds were just lucky and the true skilled funds are only 1.44%. We also provide details on the expense ratio and turnover-ratio for the average fund in each segment. First observation is that the expense ratio is similar for funds on the left/right tail and close to the median 0.83% of our sample. That means that unskilled managers are able to charge similar fees to the average of the industry while successful managers cannot claim bigger fees. In terms of turnover-ratio we see that funds on both skilled and unskilled funds exhibit higher trading activity than the median of our sample 60%.

Table 3.4: **Long-term performance**. The table displays the estimated proportions of zero-alpha, skilled and unskilled funds (π^0, π^+, π^-) for the full sample. Proportions are estimated using bootstrapped p-values according to Kosowski et al. (2006). In panel 3.4b we display the impact of luck on identifying skill. The standard errors are estimated using the Newey and West (1987) methodology and lag=6.

	π^0	π^-	π^+
Proportion	86.58	12.27	1.15

(a) Proportions of skilled/unskilled funds

	Left Tail				Right Tail				
γ	0.05	0.10	0.20	0.30	0.30	0.20	0.10	0.05	γ
Signif.	7.58	13.64	21.72	25.25	14.14	10.10	6.57	3.54	Signif.
Unlucky	2.16	4.33	8.66	12.99	12.99	8.66	4.33	2.16	Lucky
Unskilled	5.41	9.31	13.06	12.27	1.15	1.44	2.24	1.37	Skilled
α_{ann}	-2.25	-2.52	-2.53	-2.38	1.50	1.49	1.58	1.46	α_{ann}
Exp.	0.94	0.83	0.84	0.84	0.83	0.78	0.85	0.78	Exp.
Turn.	85.70	106.13	105.99	111.41	129.61	105.12	97.21	97.04	Turn.

(b) Impact of luck in the left/right tails

3.5.2 Skill and luck in short-term performance

Our results so far indicate that only a small portion of U.S. corporate bond managers are able to outperform over the long-run. However, theory suggests that outperformance can still exist over the short-term. More specifically Berk and Green (2004) propose a mechanism, where investors that chase better returns allocate heavily to skilled funds, increasing their size to the point that they lose their competitive edge in generating alpha. In this section, we examine whether mutual funds can generate alpha over the short-run. To this end, we partition our data in four non-overlapping sub-samples of three years.[6] Since we are interested in the short-term performance of each fund, we treat each fund during each sub-sample as a separate fund. After pooling our data from the four sub-periods we obtain a total of 458 p-values. We then apply the FDR methodology on the pooled p-values.

In Table 3.5a we see the results from our sub-sample analysis. As we see, performance is stronger compared to the full sample analysis, indicating that

[6]We exclude funds that have less than 24 monthly observations in the three year sub-period.

fund managers can generate alpha over the short-run, More specifically, we estimate the proportions of positive, negative and zero-alpha to be 13.52%, 10.25%, 76.23%. Our results are in line with existing literature that find evidence of mutual fund short-term over-performance. Barras et al. (2010) find similar evidence examining the U.S. equity mutual funds and Huij and Derwall (2008) find performance persistence in U.S. bond funds using a dataset from 1990 to 2003. In Appendix 3.A we provide the results of the FDR for the long-term and short-term performance using asymptotic values. While proportion estimates are more optimistic the conclusion is the same, funds are able to generate alpha over the short-term but not over the long-term. Finally, given the non-normality of residuals, we consider the bootstrapped p-values to be more appropriate in drawing inference.

Table 3.5: **Short-term performance**. The table displays the estimated proportions of zero-alpha, skilled and unskilled funds (π^0, π^+, π^-) for sub-periods of three years. We treat each fund during each sub-sample as a separate fund. Proportions are estimated using bootstrapped p-values according to Kosowski et al. (2006). In panel 3.5b we display the impact of luck on identifying skill. The standard errors are estimated using the Newey and West (1987) methodology and lag=6.

	π^0	π^-	π^+
Proportion	76.23	10.25	13.52

(a) Proportions of skilled/unskilled funds

	Left Tail				Right Tail				
γ	0.05	0.10	0.20	0.30	0.30	0.20	0.10	0.05	γ
Signif.	3.49	8.08	15.50	21.62	24.89	20.74	12.45	7.21	Signif.
Unlucky	1.89	3.79	7.58	11.37	11.37	7.58	3.79	1.89	Lucky
Unskilled	1.60	4.29	7.92	10.25	13.52	13.16	8.66	5.31	Skilled
α_{ann}	-2.33	-2.90	-2.86	-2.82	3.59	3.95	4.49	4.54	α_{ann}
Exp.	0.86	0.79	0.78	0.83	0.84	0.85	0.84	0.84	Exp.
Turn.	80.06	85.27	110.93	115.19	77.21	82.33	78.90	77.73	Turn.

(b) Impact of luck in the left/right tails

Performance persistence is another method often used to test for skill in the short-term. The main argument is that if past outperformance was due to luck, it should not persist in the future, see Carhart (1997). We further examine the hypothesis that fund managers possess skill in the short-term by creating contingency tables of past/future alpha. Every year we estimate the past alpha of the funds using the last thirty-six/twelve months of data and sort the funds into four quantiles according to their past alpha. We subsequently sort each quantile of funds according to the future alpha estimated using the following twelve months of data.

Table 3.6 shows the conditional probabilities using thirty-six and twelve months to estimate the past alpha. More specifically, in cell(i, j) of Table 3.6, we show the conditional probability of achieving a subsequent ranking of quantile j, given an initial ranking of quantile i. If the past performance had no impact on future performance, we would expect the conditional probabilities to be close to 25%, i.e., any subsequent ranking is equally likely. However, in Table 3.6, we observe strong patterns of performance persistence that hold across the two different periods that we use to estimate past alpha. For example, in panel 3.6a we see that given an initial rank at the lowest percentile, there is a 38% probability of this fund remaining in the lowest quantile in the next 12 months, while only a 17% probability that it moves to the top quantile. The same pattern holds for funds with an initial rank at the highest quantile. The probability of remaining at the top quantile in the next 12 months is 35%, while the probability of moving at the lowest quantile is 17%.

Table 3.6: **Contingency table of α_i.** The table displays the conditional probability of achieving a subsequent ranking of quantile j, given an initial ranking of quantile i. In the vertical dimension we show quantile rankings of past alpha while in the horizontal dimension we show quantile rankings of future alpha.

	Low	2	3	High
Low	37.7	23.9	20.8	17.5
2	24.6	27.9	25.2	22.3
3	21.6	22.2	32.1	24.1
High	16.9	26.0	21.7	35.4

(a) 36 month formation period

	Low	2	3	High
Low	31.2	26.8	21.9	20.0
2	27.7	25.8	25.0	21.5
3	22.3	23.6	27.3	26.8
High	19.1	23.9	25.9	31.2

(b) 12 month formation period

3.5.3 Performance persistence – Out-of-sample

Although fund outperformance over the long-run remains elusive, our results indicate that fund managers can generate alpha over shorter periods. An interesting question with direct practical implications is whether investors can identify those funds that will outperform over the next period. To answer this question, we perform an out-of-sample fund selection analysis by following the methodology of Barras et al. (2010). In order to choose among funds we use the proportion of lucky funds in the portfolio at the significance level γ:

$$\widehat{\mathrm{FDR}}_{\gamma}^{+} = \frac{\hat{F}_{\gamma}^{+}}{\hat{S}_{\gamma}^{+}} = \frac{\hat{\pi}_0 \times \gamma/2}{\hat{S}_{\gamma}^{+}} , \tag{3.13}$$

where $\mathrm{FDR}_{\gamma}^{+}$ is the proportion of lucky funds and $\hat{S}_{\gamma}^{+}$ is the proportion of funds with positive and significant t-statistic. Using the $\widehat{\mathrm{FDR}}_{\gamma}^{+}$ we can set a simple portfolio formation rule. First, we choose a $\widehat{\mathrm{FDR}}_{\gamma}^{+}$ target, which reflects how conservative we want to be concerning lucky funds. For example, setting $\widehat{\mathrm{FDR}}_{\gamma}^{+} = 10\%$, we are willing to tolerate 10% of lucky funds in our portfolio. It follows that the higher the target, the higher the proportion of lucky funds we are willing to tolerate. The portfolio formation rule proceeds as follows. Each year we estimate the alpha p-values of all existing funds using the previous 3-year period. Then we estimate $\widehat{\mathrm{FDR}}_{\gamma}^{+}$ from equation (3.13) for a grid of γ values ($\gamma = 0.01, 0.05, \ldots, 0.50$) and choose γ such that $\widehat{\mathrm{FDR}}_{\gamma}^{+}$ is closer to the target. We select funds with positive alphas and p-values smaller than γ and form an equally weighted portfolio that we hold for a year. If funds cease to exist within the holding period, we reallocate to the remaining selected funds equally. We repeat this process every year until the end of our sample.

In Table 3.7 we report the results of our selection. As we see, the FDR strategy is able to deliver economically and statistically significant alpha when we set a reasonably conservative target of lucky funds between 10% and 20%. More specifically, allowing for up to 10% of lucky funds, our strategy generates 1.75% and 1.69% of annualized alpha and is significant at a 5% level. As we increase the proportion of lucky funds allowed, the strategy's performance reduces as we go further out to the right tail of the distribution. We also report the results of a simple strategy that invests every year in the funds with alpha above/below a certain percentile, similar to Huij and Derwall (2008). The results are similar to our FDR selection strategy and confirm the hypothesis of performance persistence and skill over the short-term.[7] However, we observe that the FDR method cannot meaningfully outperform the naive strategy of portfolio formation according to past alpha. Barras et al. (2010) report similar findings for U.S. equity funds.

[7] Performance persistence is another method used to evaluate skill in the fund industry. The main argument is that if past outperformance was due to luck, it shouldn't persist in the future, see Carhart (1997).

We acknowledge that our results can put into question the effectiveness of the FDR approach in adjusting for luck. For that reason, we discuss in detail these findings. We think there are two main reasons for the inability of the FDR strategy to outperform. First, as we see from Table 3.5a, skilled funds are concentrated at the extreme right tail, with 65% of skilled funds at $\gamma = 10\%$. Second, we see that funds at the extreme rights tail also have higher alphas (4.49% for $\gamma = 10\%$ vs. 3.59% for $\gamma = 30\%$). It follows that by simply focusing on an extreme percentile like 90% we can capture a meaningful proportion of skilled funds while the benefits of identifying skilled funds further on the left are diminishing given the lower alpha and higher proportion of unskilled funds. We note that the purpose of the FDR methodology is to provide a conservative estimate of the proportions of skilled/unskilled managers and not to maximize out-of-sample performance. We further acknowledge that the precision of the FDR method can also influence the performance of the out-of-sample strategy. To this end, we evaluate the overall precision of the FDR in a detailed simulation study in the following section.

Table 3.7: **Out-of-sample performance**. In this table we display the out-of-sample performance of the FDR selection strategy. We also display a portfolio formation strategy based on a percentile rule on the estimated $\hat{\alpha}_i$. The standard errors are estimated using the Newey and West (1987) methodology and lag=6. Number of funds=198, T=158. We use *, ** and *** to indicate significance at the 10%, 5%, and 1% levels. Sample from Jan-2007 to Dec-2017, T=120.

	Mean	Std	SR	α_{ann}	t-stat
FDR10	0.56	1.25	0.45	1.75**	2.03
FDR20	0.54	1.26	0.43	1.69*	1.95
FDR40	0.51	1.26	0.40	1.04	1.10
FDR60	0.47	1.26	0.37	1.00	0.98
Perc90	0.49	1.29	0.38	2.09**	1.99
Perc75	0.53	1.31	0.41	1.90*	1.92
Perc10	0.44	1.46	0.30	-1.13**	-2.01
Perc25	0.49	1.45	0.34	-0.85*	-1.76

3.6 FDR simulation

In their critique of the FDR applicability for fund performance evaluation, Andrikogiannopoulou and Papakonstantinou (2019) argue that the simulation in the influential study of Barras et al. (2010) overstates the accuracy of FDR.

In particular, the authors raise two main points. First, they argue that the assumption that the number of months (T_i) available for each fund is equal to the total months observed in the sample (T_{max}) overstates the typical time-series length of fund returns. Second, Barras et al. (2010) use large values for the true alphas, assuming smaller alphas the FDR is inaccurate due to low signal-to-noise ratio. In their response Barras et al. (2019) argue that the size of alpha and fund volatility are related, i.e., funds with higher alpha usually exhibit higher volatility and vice-versa. Reducing the size of alpha and keeping the assumed volatility of funds constant is implicitly reducing the signal-to-noise ratio. To this end, they propose a simple approach to calibrate the two parameters according to the empirical dataset. In this section, we perform a detailed simulation study to evaluate the accuracy of the FDR procedure in our dataset, incorporating the insights of both papers. First, we generate returns according to the linear model:

$$R_{it} = \alpha_i + \beta_{1,i}\mathrm{MKT}_t + \beta_{2,i}\mathrm{TERM}_t + \beta_{3,i}\mathrm{DRF}_t + \beta_{4,i}\mathrm{LIQ}_t + \epsilon_{i,t} \tag{3.14}$$

For the distribution of the residuals, we consider two distinct cases. In the first case, we assume that the errors are homogeneous, homoscedastic, and cross-sectionally independent distributed according to $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$, with σ_ϵ^2 being the average variance across all funds in our sample. While this assumption is likely to be unrealistic, it allows us to compare results across the different simulation studies. In the second case, we consider a more complex distribution set-up for the residuals where we allow for cross-sectional dependence. We generate alphas from a discrete distribution with three point masses $(\delta^0, \delta^-, \delta^+)$, representing the zero, negative and positive alpha categories of funds. The exact distribution of alpha can be seen below:

$$\alpha \sim \pi^0\delta^0 + \pi^-\delta^- + \pi^+\delta^+ \ . \tag{3.15}$$

We initially calibrate the proportion of funds (π^0, π^-, π^+) according to our sample to 90%, 9% and 1% respectively. To evaluate the impact of proportions on the bias, we vary π^0 by 10%, keeping the ratio $\frac{\pi^-}{\pi^+} = 9$ constant. We calibrate the remaining simulation parameters in the following way:

- **Length of time-series:** We assume a balanced panel and use the median $T_{med} = 116$ as the length parameter instead of $T_{max} = 158$.
- **Residual volatility:** We use the median volatility across all funds in order to reduce the impact of outliers, similar to Barras et al. (2019).

- **Relation between fund parameters:** We create three bins that correspond to annualized alpha of $\alpha^{ann} = \{1\%, 1.5\%, 2\%\}$. We choose alphas that are conservative in terms of size but remain economically significant, representing 20%, 30%, and 40% of the annualized return of our sample. For each bin, we identify funds that fall within $\alpha_{ann} \pm 0.25\%$ and estimate the median of the residual volatility across those funds (see Table 3.8).

Table 3.8: **Simulation parameters**. This table displays the median residual volatility for different values of the true alpha α_{ann}. For each bin we identify funds that have estimated alpha $\hat{\alpha}_{ann}$that falls within $\alpha_{ann} \pm 0.25\%$ and estimate the median of the residual volatility across those funds.

α_{ann}	1	1.50	2.00
σ_{ϵ}	0.70	0.89	0.96

For each value of α_{ann} and σ_{ϵ}, we generate artificial fund returns $\tilde{R}_{it}$ from equations (3.14) and (3.15), assuming a balanced panel of returns with length T_{med}. Factor returns and betas are drawn from a normal distribution with parameters equal to the sample counterparts. We then estimate the proportions of zero-alpha, negative and positive alpha funds. We repeat this process for B=1,000 times. Finally, to measure the bias of the FDR, we calculate the miss-classification probability $\delta(\lambda)$:

$$\delta(\lambda) = \frac{\pi_A - \mathrm{E}[\hat{\pi}_A(\lambda)]}{\pi_A} \tag{3.16}$$

where π_A is the proportion of non-zero alpha funds in population and $\mathrm{E}[\hat{\pi}_A(\lambda)]$ is the average of the simulation estimates. For example a value of 20% implies that there is a 20% probability of non-zero alpha funds being miss-classified as zero-alpha.

3.6.1 i.i.d errors

In Table 3.9 we report the simulation results for the different combinations of parameters. First, we observe that the FDR method appears to be conservative in estimating non-zero alpha funds. On average, we see that the proportion of zero-alpha funds is higher than the true proportion in population and the proportion of non-zero alpha funds tends to be smaller. These results are in line with the findings of Andrikogiannopoulou and Papakonstantinou (2019) and the objective of the FDR to provide a strong control for Type-I error. The miss-classification probability $\delta(\lambda)$ is higher for smaller values of α_{ann} and π_0. The size of α_{ann} is the most important parameter with $\delta(\lambda)$ ranging from 39.2% to 49.9% for $\alpha_{ann} = 1$, 31.6% to 38.2% for $\alpha_{ann} = 1.5$ and 4.9% to 12.1% for $\alpha_{ann} = 2$.

Comparing simulation results for different datasets is not straightforward. However, we think it is reasonable to compare the bias of our conservative case $\alpha_{ann} = 1$ and the base case $\alpha_{ann} = 2$ for equity mutual funds reported in Andrikogiannopoulou and Papakonstantinou (2019) and Barras et al. (2019). More specifically, Andrikogiannopoulou and Papakonstantinou (2019) report a 65% miss-classification probability in that case while Barras et al. (2019) report 48%. Our simulation results are less pessimistic than Andrikogiannopoulou and Papakonstantinou (2019) and closer to the recent response of Barras et al. (2019). Moreover, it is encouraging that for $\alpha = 1.5$ (the alpha values we observe in sample), the miss-classification probability declines to a range of 31.6% to 38.2%. Overall, our simulation results are comparable to those for equity mutual funds. Thus, we conclude that the FDR can be successfully applied in the cross-section of corporate bond mutual funds.

3.6.2 Cross-sectionally dependent errors

We further evaluate the accuracy of the FDR in our sample for the case that the errors are cross-sectionally dependent. We follow Barras et al. (2010) and use the residuals factor specification proposed by Jones and Shanken (2005). The residuals of each fund i are specified according to:

$$\epsilon_{it} = \delta G_t + \delta G_t^- I_{\alpha_i<0} + \delta G_t^+ I_{\alpha_i>0} + \xi_{it} \, . \tag{3.17}$$

We assume that the errors of all the funds have exposure to a common factor G_t, while only negative/positive alpha funds load on factors G_t^- and G_t^+, respectively; δ is the common loading of the funds on these factors. Finally, ξ_{it} is the independent part of the error and is normally distributed according to $\xi_{it} \sim N(0, \sigma_\xi^2)$. We further assume that G_t, G_t^- and G_t^+ are orthogonal to each other and follow a normal distribution $N(0, \sigma_G)$. We calibrate parameters σ_G and δ from our sample. We set $\sigma_G = 0.024$, which is equal to the average standard deviation of DRF and LIQ. We set $\delta = 0.14$, equal to the average beta of the funds on DRF and LIQ. We fix σ_ξ, such that the standard deviation of residuals $\sigma_{\epsilon_{it}}$ of each fund equals to the median residual volatility described in Table 3.8.

In Table 3.10 we report the simulation results for the residual factor dependence specification. The miss-classification probability $\delta(\lambda)$ ranges from 39.4% to 49.1% for $\alpha_{ann} = 1$, 26.8% to 38.8% for $\alpha_{ann} = 1.5$ and 3.8% to 12.9% for $\alpha_{ann} = 2$. As we see, our results are comparable to the baseline case of i.i.d errors. Barras et al. (2010) come to a similar conclusion, finding no impact on the performance of the FDR for the case of cross-sectionally dependent errors. Overall, we conclude that the assumptions about the true value of alpha and proportions of funds have the biggest impact on the performance of the FDR. Our simulation study uses a wide range of economically sensible parameters for the true alpha and proportions of funds. We find that the FDR performs adequately even under conservative assumptions.

Table 3.9: **Simulation results (i.i.d errors)**. The table displays the average value of the estimated proportions of funds with zero, negative and positive alpha (π^0, π^+, π^-) across the different simulated scenarios. Errors are assumed to iid. At the bottom of each sub-table we report the miss-classification probability $\delta(\lambda)_{\text{sim}}$. T=116, B=1000.

α_{ann}	1.0	1.5	2.0
$\pi^0 = 90$	93.1	92.6	90.2
$\pi^- = 9$	5.7	6.3	8.6
$\pi^+ = 1$	0.3	0.5	0.9
$\delta(\lambda)_{\text{sim}}$	39.2	31.6	4.9

α_{ann}	1.0	1.5	2.0
$\pi^0 = 80$	88.5	86.8	81.6
$\pi^- = 18$	10.4	12.1	16.3
$\pi^+ = 2$	0.3	0.5	1.6
$\delta(\lambda)_{\text{sim}}$	46.2	36.6	10.5

α_{ann}	1.0	1.5	2.0
$\pi^0 = 70$	83.5	80.1	73.1
$\pi^- = 27$	15.5	18.3	24.3
$\pi^+ = 3$	0.2	0.8	2.2
$\delta(\lambda)_{\text{sim}}$	47.7	36.3	11.6

α_{ann}	1.0	1.5	2.0
$\pi^0 = 60$	78.9	74.4	64.2
$\pi^- = 36$	20.1	23.9	32.2
$\pi^+ = 4$	0.0	0.8	2.9
$\delta(\lambda)_{\text{sim}}$	49.9	38.2	12.1

Table 3.10: **Simulation results (cross-sectionally correlated errors)**. The table displays the average value of the estimated proportions of funds with zero, negative and positive alpha (π^0, π^+, π^-) across the different simulated scenarios. Errors are assumed to be cross-sectionally correlated according to equation (3.17). At the bottom of each sub-table we report the miss-classification probability $\delta(\lambda)_{\text{sim}}$. T=116, B=1000.

α_{ann}	1.0	1.5	2.0
$\pi^0 = 90$	93.0	92.0	90.0
$\pi^- = 9$	5.6	6.6	8.6
$\pi^+ = 1$	0.5	0.7	1.0
$\delta(\lambda)_{\text{sim}}$	39.4	26.8	3.8

α_{ann}	1.0	1.5	2.0
$\pi^0 = 80$	88.3	86.5	81.5
$\pi^- = 18$	10.5	12.3	16.5
$\pi^+ = 2$	0.3	0.7	1.6
$\delta(\lambda)_{\text{sim}}$	45.9	35.4	9.7

α_{ann}	1.0	1.5	2.0
$\pi^0 = 70$	83.3	79.8	73.3
$\pi^- = 27$	15.5	18.4	24.3
$\pi^+ = 3$	0.3	0.8	2.0
$\delta(\lambda)_{\text{sim}}$	47.6	35.9	12.1

α_{ann}	1.0	1.5	2.0
$\pi^0 = 60$	78.6	74.4	64.8
$\pi^- = 36$	20.3	23.7	32.0
$\pi^+ = 4$	0.1	0.7	2.8
$\delta(\lambda)_{\text{sim}}$	49.1	38.8	12.9

3.7 Conclusion

In this paper, we analyze the performance of U.S. corporate mutual funds and attempt to distinguish skill from luck for individual fund managers. We utilize the FDR methodology of Barras et al. (2010) and a reduced version of the asset pricing model introduced by Bai et al. (2019). We first analyze the ability of fund managers to outperform over the long-run. To this end, we apply the FDR approach to the full sample and find that 1.15%, 12.27%, and 86.58% of the funds generate positive, negative, and zero-alpha. Thus, we conclude that alpha generation is scarce over the long-run, and most funds generate zero-alpha. We continue our analysis by evaluating the performance of funds over the short-run. According to the theoretical work of Berk and Green (2004), funds might be able to outperform over the short-run before investors reduce their competitive advantage by increasing inflows. We find that performance is stronger compared to the full sample analysis, indicating that fund managers can generate alpha over the short-run. We estimate the proportions of positive, negative, and zero-alpha to be 13.52%, 10.25%, 75.80%. Motivated by the fact that corporate bond funds can outperform over the short-run, we examine whether investors can use past performance to identify talented managers and create profitable strategies. We create a selection strategy based on the FDR methodology and show that we can identify managers that possess skill. Our strategy generates economically and statistically significant annualized alpha of 1.75%. Our results are economically meaningful for investors suggesting that dynamic and active manager selection pays off. Finally, given the recent critique of Andrikogiannopoulou and Papakonstantinou (2019) on the bias of the FDR methodology, we perform a detailed simulation study to evaluate the performance of FDR in our dataset. We use the median size of our sample T_{med} and calibrate the alpha and residual volatility parameter from our data in line with Barras et al. (2019). While we confirm previous findings of bias in FDR estimates, our simulation results are less pessimistic than Andrikogiannopoulou and Papakonstantinou (2019) and closer to the recent response of Barras et al. (2019). Overall, our simulation results are comparable to those for equity mutual funds. Thus, we conclude that the FDR can be successfully applied in the cross-section of corporate bond mutual funds.

References

Andrikogiannopoulou, A., & Papakonstantinou, F. (2019). Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power? *The Journal of Finance*, *74*(5), 2667–2688.

Ayadi, M. A., & Kryzanowski, L. (2011). Fixed-income fund performance: Role of luck and ability in tail membership. *Journal of Empirical Finance*, *18*(3), 379–392.

Bai, J., Bali, T. G., & Wen, Q. (2019). Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, *131*(3), 619–642.

Bao, J., Pan, J., & Wang, J. (2011). The Illiquidity of Corporate Bonds. *The Journal of Finance*, *66*(3), 911–946.

Barras, L., Scaillet, O., & Wermers, R. (2010). False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas. *The Journal of Finance*, *65*(1), 179–216.

Barras, L., Scaillet, O., & Wermers, R. (2019). Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power? A Reply. *The Journal of Finance, Forthcoming*.

Berk, J. B., & Green, R. C. (2004). Mutual Fund Flows and Performance in Rational Markets. *Journal of Political Economy*, *112*(6), 1269–1295.

Bessembinder, H., Kahle, K. M., Maxwell, W. F., & Xu, D. (2009). Measuring Abnormal Bond Performance. *The Review of Financial Studies*, *22*(10), 4219–4258.

Blake, C. R., Elton, E. J., & Gruber, M. J. (1993). The Performance of Bond Mutual Funds. *The Journal of Business*, *66*(3), 371–403.

Brooks, J., Gould, T., & Richardson, S. (2020). Active Fixed Income Illusions. *The Journal of Fixed Income*, *29*(4), 5–19.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, *52*(1), 57–82.

Christiansen, C., Grønborg, N. S., & Nielsen, O. L. (2020). Mutual fund selection for realistically short samples. *Journal of Empirical Finance*, *55*, 218–240.

Cici, G., & Gibson, S. (2012). The Performance of Corporate Bond Mutual Funds: Evidence Based on Security-Level Holdings. *The Journal of Financial and Quantitative Analysis*, *47*(1), 159–178.

Cremers, K. M., Fulkerson, J. A., & Riley, T. B. (2019). Challenging the Conventional Wisdom on Active Management: A Review of the Past 20 Years of Academic Literature on Actively Managed Mutual Funds. *Financial Analysts Journal*, *75*(4), 8–35.

Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods* (Vol. 5). New York: Oxford University Press.

Dick-Nielsen, J. (2009). Liquidity Biases in TRACE. *The Journal of Fixed Income*, *19*(2), 43–55.

Dick-Nielsen, J. (2014). *How to Clean Enhanced TRACE Data* (SSRN Working Paper No. 2337908).

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2010). Luck versus Skill in the Cross-Section of Mutual Fund Returns. *The Journal of Finance*, *65*(5), 1915–1947.

Ferson, W., Henry, T. R., & Kisgen, D. J. (2006). Evaluating Government Bond Fund Performance with Stochastic Discount Factors. *The Review of Financial Studies*, *19*(2), 423–455.

Harvey, C. R., & Liu, Y. (2018). Detecting Repeatable Performance. *The Review of Financial Studies*, *31*(7), 2499–2552.

Huij, J., & Derwall, J. (2008). "Hot Hands" in bond funds. *Journal of Banking & Finance*, *32*(4), 559–572.

Ivashchenko, A. (2019). *Corporate Bond Price Reversals* (SSRN Working Paper No. 3473739).

Jones, C. S., & Shanken, J. (2005). Mutual fund performance with learning across funds. *Journal of Financial Economics*, *78*(3), 507–552.

Konstantinov, G. S., & Fabozzi, F. J. (2021). Towards a dead end? EMU bond market exposure and manager performance. *Journal of International Money and Finance*, *116*, 102433.

Kosowski, R., Timmermann, A., Wermers, R., & White, H. (2006). Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap

Analysis. *The Journal of Finance*, *61*(6), 2551–2595.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708.

Pástor, Ľ., Stambaugh, R. F., & Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, *116*(1), 23–45.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(3), 479–498.

3.A Results with asymptotic p-values

Table 3.11: **Long-term performance**. The table displays the estimated proportions of zero-alpha, skilled and unskilled funds (π^0, π^+, π^-) for the full sample. Proportions are estimated using asymptotic p-values. In panel 3.11b we display the impact of luck on identifying skill. The standard errors are estimated using the Newey and West (1987) methodology and lag=6.

	π^0	π^-	π^+
Proportion	77.92	16.59	5.48

(a) Proportions-Asymptotic p-values

	Left Tail				Right Tail				
γ	0.05	0.10	0.20	0.30	0.30	0.20	0.10	0.05	γ
Signif.	10.10	17.68	22.73	28.28	17.17	13.64	7.07	6.57	Signif.
Unlucky	1.95	3.90	7.79	11.69	11.69	7.79	3.90	1.95	Lucky
Unskilled	8.15	13.78	14.94	16.59	5.48	5.84	3.17	4.62	Skilled
α_{ann}	-2.34	-2.56	-2.41	-2.26	1.47	1.56	1.63	1.65	α_{ann}
Exp.	0.86	0.82	0.79	0.85	0.81	0.83	0.84	0.80	Exp.
Turn.	95.33	109.70	109.85	124.03	121.16	133.23	116.45	118.30	Turn.

(b) Impact of luck-Asymptotic p-values

Table 3.12: **Short-term performance**. The table displays the estimated proportions of zero-alpha, skilled and unskilled funds (π^0, π^+, π^-) for sub-periods of three years. We treat each fund during each sub-sample as a separate fund. Proportions are estimated using asymptotic p-values. In panel 3.12b we display the impact of luck on identifying skill. The standard errors are estimated using the Newey and West (1987) methodology and lag=6.

	π^0	π^-	π^+
Proportion	64.88	15.60	19.53

(a) Proportions-Asymptotic p-values

	Left Tail				Right Tail				
γ	0.05	0.10	0.20	0.30	0.30	0.20	0.10	0.05	γ
Signif.	9.83	14.63	19.43	25.33	29.26	25.11	21.18	15.72	Signif.
Unlucky	1.62	3.24	6.49	9.73	9.73	6.49	3.24	1.62	Lucky
Unskilled	8.20	11.38	12.94	15.60	19.53	18.62	17.94	14.10	Skilled
α_{ann}	-2.80	-2.97	-2.75	-2.77	3.26	3.55	3.94	4.37	α_{ann}
Exp.	0.78	0.79	0.80	0.85	0.85	0.84	0.84	0.84	Exp.
Turn.	92.13	114.37	117.07	114.02	75.28	77.68	81.59	88.49	Turn.

(b) Impact of luck-Asymptotic p-values

3.B Factor data

We use OTC bond transaction data which is available directly through Financial Industry Regulatory Authority's (FINRA) Trade Reporting and Compliance Engine (TRACE). TRACE was introduced by the National Association of Securities Dealers (NASD) in 2002, to improve transparency in the OTC corporate bond market. From February 2005 onwards, 99 percent of all TRACE-eligible bond transactions are covered on an intra-day basis. The information covered in the TRACE database is listed by transaction, and key variables include transaction date, time, price, and traded volume. Therefore, TRACE is the most comprehensive source of pricing information when it comes to research questions concerning the U.S. corporate bond market. We use the enhanced version of the TRACE database, which has no volume cap on reported trades and thus captures a broader range of the U.S. corporate bond market. Our data extend from July 2002 to December 2018. Since the TRACE data is dealer reported, errors can happen. Instead of correcting the data in the database directly, trade messages are appended, indicating either cancellations, corrections, or reversals. Trades can also be double counted in the TRACE system because various parties can report the same trade. We follow the steps outlined in Dick-Nielsen (2009) and Dick-Nielsen (2014) to take care of the cancellation, correction, reversal, and double counting issues.

In our effort to clean the TRACE data, we follow the cleaning steps outlined in Bessembinder, Kahle, Maxwell, and Xu (2009) and Bai et al. (2019). In particular, we remove transaction records with a trading volume of less than 10,000 USD, are labeled as when-issued, locked-in, have special sales conditions, have more than a two-day settlement, and are flagged as equity-linked notes. In order to minimize the effect of bid-ask spreads in prices, we calculate the daily clean price as the trading volume-weighted average of intra-day prices. We continue by following Bai et al. (2019) in removing trade records that feature a transaction price under 5 or above 1,000 USD. This step implicitly removes some defaulted bonds. However, this is a very low threshold in our view since many defaulted bonds tend to trade above 5 USD. Including defaulted bonds in our study can create biased results for three main reasons: 1) defaulted bonds typically do not accrue interest, 2) liquidity is typically very low after a default, and 3) actual recovery is deal-specific and hard to estimate. For these reasons and in contrast to previous papers such as Bai et al. (2019), we choose to control for defaulted bonds directly. We incorporate the issue-level default data information from the Altman[8] database and exclude all future bond observations post a given default date. In the initial results, we observed an overly strong reversal factor. After checking data manually, we found that extreme day-to-day reversals exist due to prices of small trades and thus is a type of an outlier. For this reason, we implemented a reversal rule, where we exclude trade records that featured a daily reversal of more than 10 USD in absolute terms. E.g. if a bond trades at 100 on Monday, drops to below 90 on Tuesday, and then features again a price of over 100 on Wednesday, we exclude the trade record for Tuesday for that particular bond. We then aggregate the daily bond price data to the monthly frequency. We follow Bai et al. (2019) in only considering a bond's monthly end price as a non-missing value if it falls within the last seven weekdays of that particular month. The monthly traded volume for each bond is generated by first aggregating it to the daily level by taking the intra-day sum for all days and then taking the sum over all days in a particular month.

We then match the bond CRSP link database to get the corresponding firm identifier for each bond and match the corresponding equity and financial variables from CRSP and Compustat. We retrieve bond issue information such as the maturity date, coupon rate, coupon payment frequency, etc., from the Mergent FISD database. A key metric of a bond is its rating and how it changes over time. We access the historical rating history of each bond through Mergent FISD's historical bond rating database. It includes ratings from Standard & Poor's, Moody's, and Fitch. We follow the industry convention and form composite ratings by applying the following methodology: if three outstanding ratings are available, we choose the middle rating; if two ratings are outstanding, we choose the more conservative one, and if only one rating is outstanding, we choose the one that is available.

We use information from the Mergent FISD database in order to further filter our data. We exclude bonds that are not listed or traded in the U.S. public market. That includes private placements, 144A bonds, bonds that do not trade in U.S. dollars, and issuers not located in the U.S. We further remove structured notes, mortgage-backed securities, asset-backed securities, agency-backed bonds, and convertible bonds. We only keep bonds with a fixed or a zero-coupon and exclude all bonds with a variable or floating coupon rate. In addition to excluding perpetuals (bonds without a fixed maturity), we exclude bonds with a remaining lifetime of more than 30 years since these bonds tend to be illiquid. If a bond trades close to maturity, i.e. less than one year, it

[8] Altman Kuehne NYU Salomon Center Corporate Bond Default Master Database.

is delisted from major corporate bond indices, thus reducing its liquidity. We thus remove bonds with less than one year remaining to maturity.

Afterwards, we calculate for each bond-month observation its corresponding yield-to-maturity and modified duration. Both of these variables are key metrics needed in the construction of characteristics. In order to calculate the total return of a bond, we also need to incorporate information on when coupon payments were made and what the accrued interest amounted to at each point in time for each individual bond issue. The coupon payment schedule can be backed out by going backwards from the last coupon payment by the bond's coupon payment frequency. Accrued interest is the amount of interest that accrues from one coupon payment to the next. We follow Bessembinder et al. (2009) in their calculation of a bond's total return:

$$r_{i,t} = \frac{P_{i,t} + AI_{i,t} + C_{i,t}}{P_{i,t-1} + AI_{i,t-1}} - 1, \tag{3.18}$$

where $P_{i,t}$ is the transaction price (i.e. the clean price), $AI_{i,t}$ accrued interest and $C_{i,t}$ the coupon payment if any, of bond i in month t. We denote bond i's excess return as $R_{i,t} = r_{i,t} - r_{f,t}$, where $r_{f,t}$ is the return of the one month T-Bill. As a final step in our sample setup, we adopt the same bond trading/liquidity restriction Bai et al. (2019) does. In particular, for a bond-month observation to be considered in our analysis, we require that it had a valid return (i.e. a price in the last 7 weekdays of a month) for at least 24 out of the past 36 months.

Table 3.13: **Sample setup**. In this table we describe the detailed steps we take to filter the TRACE dataset and the impact of those steps on the dimensions of our dataset.

	issues	firms	frequency	obs
starting sample, TRACE raw data	184,645		intra-day	223,961,512
- daily aggregation	173,017		daily	22,063,238
- remove trade records, price < 5, > 1000 USD	172,422		daily	22,024,034
- remove trade records, price reversal $> \|10\|$ USD	172,422		daily	22,005,801
- monthly aggregation	66,026		monthly	1,594,457
- match bond CRSP link database	48,618	3,022	monthly	1,152,419
- match bond default database	48,598	3,015	monthly	1,142,465
- match Mergent FISD issue database	25,994	1,889	monthly	738,947
- match Mergent FISD ratings database	24,002	1,775	monthly	696,258
- remove trade records, < 24 out of 36 months traded	22,955	1,744	monthly	619,498
- match bond variables	10,836	1,254	monthly	336,707

Table 3.13 shows the sample setup grouped into the main cleaning categories

described above. The sample starts with the intra-day level frequency of the raw TRACE data consisting of 184,645 bond issues and ends with a final sample size of 7,839 bond issues from 1,087 unique firms. The step with the largest loss of observations is aggregating the daily observations to the monthly level. This step reduces the amount of issues roughly by 60 percent. This is largely due to our restriction that in order to have a non-missing return in a given month, the bond needs to trade in the last 7 weekdays of two consecutive months. Our above mentioned filtering criteria after matching in the Mergent FISD issue database reduces the amount of issues covered from 48,598 to 25,994. Only few observations were lost due to non-available rating information. These in part were also due to withdrawal of ratings due to imminent maturity. A sizeable fraction of observations was lost at the end due to the necessity of a 36 month rolling window in calculating specific key bond characteristics such as the value at risk or expected shortfall metrics from Bai et al. (2019).

Table 3.14 shows the main summary statistics for our sample of corporate bonds. The mean, median, standard deviation and percentiles were calculated by pooling all bond-month observations. On average, corporate bonds exhibited a total return of 0.54 percent in a given month. 75 percent of observations can be categorized as investment grade. Bonds on average were issued 6.52 years ago and have a remaining 8.53 years until maturity. The average coupon and yield to maturity are 5.70 and 4.64 percent, respectively. 650.37 million USD is the average amount issued.

Table 3.14: **Summary statistics (Not CRSP/Compustat matched)**. In this table we display the summary statistics of our sample when we do not match with the CRSP/Compustat databases. Data from October-2005 to December-2018.

					Percentiles					
	N	Mean	Median	SD	1st	5th	25th	75th	95th	99th
Bond total return (percent)	336,707	0.54	0.34	6.00	-11.78	-3.46	-0.38	1.30	4.58	12.52
Rating (1-21, 1=AAA, IG $\leq$ 10)	336,707	8.86	8	3.77	1	4	6	10	16	20
Investment grade (1=IG, 0=HY)	336,707	0.75	-	-	-	-	-	-	-	-
Time to maturity (years)	336,707	8.81	5.50	7.88	1.08	1.42	3.17	13.25	25.50	26.83
Age (years)	336,707	6.84	5.58	4.05	3.08	3.25	4.08	7.92	16.17	21.83
Coupon (percent)	336,707	5.91	6.00	1.64	1.85	3.10	5.00	6.95	8.50	9.88
Yield to maturity (percent)	336,707	5.46	4.51	7.38	0.71	1.32	2.85	6.16	10.99	29.49
Modified duration (years)	336,707	5.83	4.60	3.98	1.08	1.36	2.79	8.38	13.72	15.52
Offering amount (million USD)	336,707	619.50	500.00	618.78	17.28	30.00	250.00	750.00	1998.33	3000.00

Curriculum Vitae

Education

- Bachelor in Economics, University of St.Gallen, 2014.
- Master in Quantitative Economics and Finance, University of St.Gallen, 2016.
- Ph.D. in Economics and Finance, University of St.Gallen, 2022.

Work experience

- Quantitative research analyst, Wellershoff & Partners Ltd., 2015–2017.
- Research assistant, Faculty of Mathematics and Statistics, University of St.Gallen, 2017–2022.

Teaching

- Statistics – Exercises, Bachelor in Economics, University of St.Gallen, 2017–2018.
- Data Analytics I: Statistics – Exercises, Bachelor in Economics, University of St.Gallen, 2019–2021.
- Data Analytics I: Statistics – Tutorials, Bachelor in Economics, University of St.Gallen, 2017–2019.
- Workshop Fundamentals of Data Science, Certificate in Data Science Fundamentals, Bachelor: Additional Qualification Programmes, University of St.Gallen, 2018–2021.

Presentations

- Conference on Computational and Financial Econometrics, University of London, 2019.