Corporate Bond Factors: Replication Failures and a New Framework

Jens Dick-Nielsen, Peter Feldhütter, Lasse Heje Pedersen, and Christian Stolborg*

First version: January 12, 2023. This version: December 9, 2023

Abstract

We demonstrate that the literature on corporate bond factors suffers from replication failures, inconsistent methodological choices, and the lack of a common error-free dataset. Going beyond identifying this replication crisis, we create a clean database of corporate bond returns where outliers are analyzed individually and propose a robust factor construction. Using this framework, we show that most, but not all, factors fail to replicate. Further, while traditional factors are constructed from individual bonds, we create representative firm-level bonds, showing which bond signals work at the firm-level. Lastly, we show that a number of equity signals work for corporate bonds. In summary, most factors fail, but so does the CAPM for corporate bonds.

Keywords: asset pricing, factors, data mining, replication, multiple testing, external validity

JEL Codes: G11, G12, G14, G32, G4, C5

*Dick-Nielsen (jdn.fi@cbs.dk), Feldhütter (pf.fi@cbs.dk), and Stolborg (cst.fi@cbs.dk) are at Copenhagen Business School. Pedersen is at AQR Capital Management, Copenhagen Business School, and CEPR; www.lhpedersen.com. We are grateful for helpful comments from Dion Bongaerts, Alex Dickerson, Andrea Frazzini, Antti Ilmanen, Scott Richardson, Michael Schwert, and participants at seminars at Copenhagen Business School and the BIGFI Conference, March 2023. We thank Frederik Andersen, Felix Lunden, Christoffer Algreen Nielsen, and Marcus Piil Pedersen for excellent research assistance. We gratefully acknowledge support from the Center for Big Data in Finance (Grant no. DNRF167). AQR Capital Management is a global investment management firm, which may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR.

Many fields of research have been accused of facing a credibility crisis, starting with medicine and more recently the finance literature on equity factors.[1] We examine the replicability of the rapidly growing literature on corporate bond factors. In other words, which factors drive expected corporate bond returns and how credible is the literature that studies this question?

In short, we find that most of the corporate bond factors in the literature fail to replicate, in large part due to problems with the underlying data, combined with varying and non-robust ways of dealing with these data errors. We present a new framework of clean data and robust factor construction, finding a minority of significant corporate bond factors. We also present a way to aggregate each firm's many bonds into a single time series of representative bond returns for each firm. Using these firm-level corporate bond returns as test assets, we study factor returns based on the existing corporate-bond signals as well as signals from the literature on equity factors. We intend to make our clean corporate bond returns, firm-level returns, factor returns, and code available to researchers.

As a way of background, Harvey et al. (2016) and Hou et al. (2020) propose that the finance literature on equity factors faces a replication crisis. However, Chen and Zimmermann (2020) show a 98% reproduction rate, meaning that most factors remain significant when they use the same method and data. This high reproduction rate is encouraging, but does not settle the issue, since part of the criticism is that the original papers use tailored non-robust methods and did not account for multiple testing. Addressing these issues, Jensen et al. (2022) document a high degree of scientific replication (replicating factors with a consistent robust methodology that accounts for multiple testing) and external validity (out-of-sample evidence across time and countries).

This high replication rate of the equity factor literature may be due to its solid foundation based on common underlying data sets (CRSP and Compustat) with a relatively low incidence of data errors and biases. In contrast, the literature on corporate bond factors stands on a more shaky foundation as this asset class does not have a common error-free data set. In other words, even before each corporate bond paper chooses a factor-contruction method

[1]The areas with a debate on replication crises (or credibility crises) include medicine (Ioannidis, 2005), psychology (Nosek et al., 2012), management (Bettis, 2012), experimental economics (Maniadis et al., 2017), and other fields of research.

(as with equity factors), the paper first decides how to "clean" the data. Some papers use raw data from TRACE, others add different filters, and most researchers use winsorized data from WRDS, but all these methods involve errors.[2] This data issue may seem surprising given that the corporate bond asset class is of similar size and importance as equities, but corporate bond data are more complex for several reasons: (i) whereas each company has a single stock, a company can have more than 100 bonds outstanding, (ii) the set of outstanding bonds changes over time, and (iii) some bond series are highly illiquid and the raw data have numerous erroneous and real outliers. Addressing these data challenges with our new framework, our findings are as follows.

Reproduction: Data-cleaning method matters. Our first finding is that basic data-cleaning choices lead to significant replication problems in the corporate bond literature. To analyze the effects of these data choices, we first reproduce the most cited factors in the literature with each original paper's own data choices. We find that all factor risk premia have a point estimate with the same sign, but only one of the risk premia remains significant (i.e., only one positive reproduction). Then, to study robustness, we compute the return of each factor with the data choices from the *other* papers. Done this way, none of the factor risk premia are significant with the choices used in the other papers, a form of failure of scientific replication. In fact, the point estimate of the average excess return even changes sign in some cases.

Clean data matters. We are not just interested in examining the credibility of the literature, we are also searching for the truth about credit market returns. Therefore, we seek to create a relatively clean data set of corporate bond returns. Rather than making arbitrary choices, we first apply filters that eliminate a range of known errors, and then analyze all the most extreme remaining outliers "by hand," eliminating errors and retaining extreme returns that represent real economic events. Furthermore, we include returns around defaults. Using this clean data, we find that most corporate bond factors fail to replicate, even with each paper's own factor construction methodology. At a more basic level, we find

[2] For example, the WRDS Bond Database winsorizes returns greater than 100% (i.e., leaving in errors, but reducing their effect, and eliminating real large economic events), includes non-standard bonds such as convertible bonds, miscalculates accrued interest (e.g., ignoring settlement dates), and use a default return that we cannot replicate. We instead check outliers individually, ensure that all bonds are standard corporate bonds, calculate accrued interest carefully, and empirically justify our default return.

that the average return of the overall corporate bond market is meaningfully different using the clean data versus some of the data cleaning methods in the literature.

Scientific replication: A robust framework. While the literature uses different data-cleaning and factor-construction methods, we are interested in examining factor returns using the same clean data set and a consistent robust factor-construction method. In particular, the different papers in the literature use factor-construction methods based on (i) equal- or value-weighting; (ii) tertile, quintile, or decile portfolios; and (iii) single- or double-sorting. We argue that a robust method is to (i) value-weighted returns for implementability and to reduce the importance of missing returns; (ii) use tertile portfolios to include a large fraction of the data; and (iii) double-sort based on each signal and three broad credit-rating groups to ensure apples-to-apples comparisons.

Specifically, each month we first sort all bonds into three ratings categories: investment grade plus, "IG+" (i.e., all ratings that start with A), investment grade minus, "IG-" (i.e., all ratings that start with BBB or Baa), and speculative grade, "SG" (i.e., all lower ratings). Within each of these three ratings categories, we sort all bonds into three groups based on the signal of interest — for example, yield-to-maturity (YTM), momentum, or bond age. We then compute value-weighted returns for each rating-by-signal group and create long-short portfolios within each rating category — for example, buying high-YTM bonds while shorting low-YTM ones within IG+. This method yields factor returns within each rating group IG+, IG-, and SG—each interesting in its own right—and our overall factor return is the average of these three factor returns.

Based on our clean data and robust factor construction method, we replicate all of the factors from the literature, and compute their alphas controlling for the overall credit market return and the overall Treasury bond return. We find that only 23% of the factors considered significant in the literature have significant alphas using our framework, as seen in the first bar in Figure 1. This finding reveals a surprisingly low scientific replication rate.

Out-of-sample evidence. While the literature is focused on the factor performance over the time period from 2002 when TRACE data is available, we also construct factor returns from 1985 using alternative data sources based on quotes. Over the pre-TRACE time period, we also find a low replication rate of 27% as seen in Figure 1. Merging the

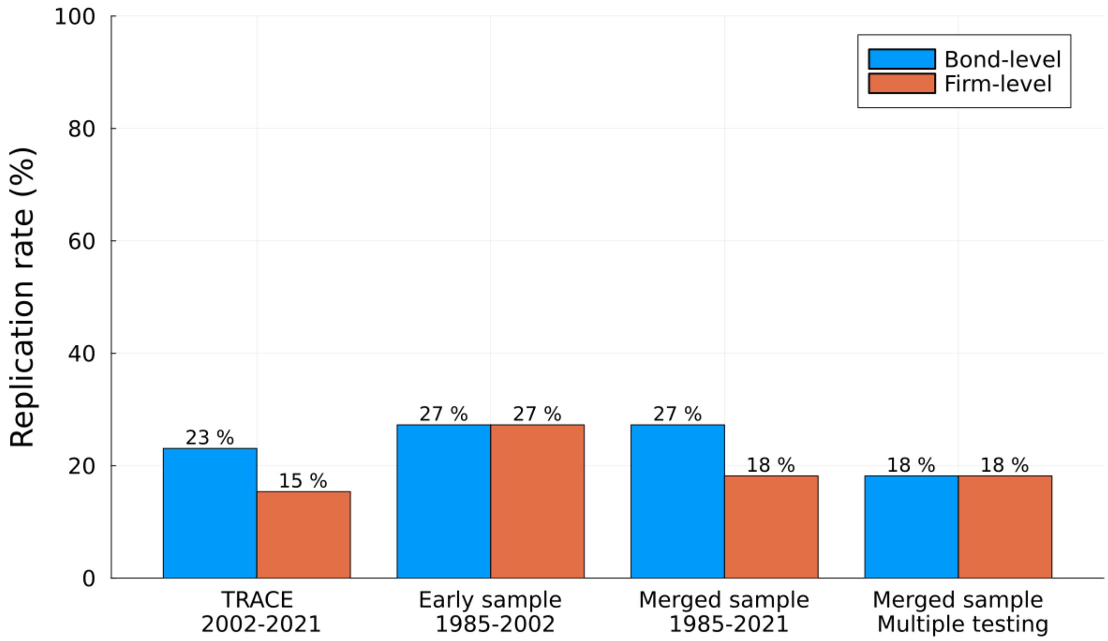

Figure 1: **Replication rates.** This figure shows the fraction of factors from the literature that are replicated, meaning that their alphas are significant and of the same sign as the original paper. The alpha is the intercept in the regressing of the factor's excess return on that of the overall credit market and the overall Treasury market.

recent and old data to get as much information as possible on factor returns, we also find a 27% replication rate, as seen in Figure 1.

Firm-level corporate bond returns. As another innovation of our framework, we create firm-level representative bond returns by value-weighting all outstanding bonds for each firm. These firm-level bond returns have several potential applications. First, having a single time series of representative bond returns for each firm makes bond data much easier to work with for other researchers, creating a broadly applicable data set that parallels that for equities. Second, the focus on individual bonds in the literature means that firms with many bonds may get over-represented in the factor construction, including how thresholds are set for the tertile portfolios. Third, having both bond-level and firm-level returns is helpful in recognizing whether factors are driven by firm-level or bond-level characteristics. These firm-level factors imply slightly lower replication rates in the TRACE and merged samples, but the same replication rate in the early sample, as seen in Figure 1.

Corporate bond factors based on equity characteristics. An old question in finance is whether the same factors price stocks and corporate bonds (e.g., Fama and French (1993)). In this spirit, Chordia et al. (2017) consider whether 10 characteristics from the literature on equity factors also work for corporate bonds. We replicate and extend this analysis, considering 153 signals from Jensen et al. (2022). Since we have one equity signal per firm, testing the predictive power for corporate bonds is straightforward using our firm-level corporate bond returns.

Multiple-testing (MT) correction. Given that we are testing the significance of each of many factors, we need to account for multiple testing problems. This issue is especially acute when we consider 153 equity signals. Indeed, even if none of the equity signals were truly significant, 5% of them (approximately 8 factors) would appear to be significant just by chance.

We use the most standard frequentist MT adjustment, Benjamini and Hochberg (1995). Such MT adjustments depend on the collection of all the tests tried by researchers, in principle both those that are published and those that are not. While knowing all tried tests is never possible, we use a comprehensive set based on all corporate bond factors at the bond level and at the firm level, as well as all factors based on equity signals (as a proxy for everything that researchers might have tried). This method implies that a factor is considered significant if its alpha has a t-statistic greater than 2.43 in absolute magnitude, rather than the regular cut-off of 1.96.

Using this cut-off based on the MT correction, fewer factors are significant. In the merged sample, 18% of the corporate bond factors are significant at the bond or firm level, as seen in Figure 1. The factors that are significant in both bond and firm merged samples with MT correction are bond age, market value, and equity momentum. Further, Duration is significant at the bond level and Amount Outstanding at the firm level. In other words, high-return bonds have historically been those that are old, small market value, high recent return of the corresponding stock, low duration, and from a small firm (i.e., a low total amount outstanding across all the firm's bonds).

Turning to the factors based on equity signals, we find that a large number of factors are significant even when adjusting for multiple testing effects. In particular, 33 factors have significantly positive returns, including factors based on various forms of equity momentum, investment, and debt issuance. Also, 16 factors have significantly negative returns, including factors based on short-term reversal (said differently, short-term equity momentum is a positive predictor of corporate bonds, just like regular equity momentum) and value factors (which could imply a poor performance of bonds from firms that the equity market views as financially distressed). The factors that predict equity and bond returns with the same sign may work because they predict the overall performance of the firm — in contrast, value factors that predict equities positively and bonds negatively may capture changes in risk or other effects that redistribute value across the capital structure.

Related literature. Our paper is related to several literatures. First, the paper is related to the broader literature on replication, as cited above. Second, the paper is clearly related to all the corporate bond factor papers that we replicate, cited in the methodology section. We do not replicate papers that go beyond simple long-short factors, such as Kelly et al. (2023) who use instrumented principal components analysis, Acharya et al. (2013), Greenwood and Hanson (2013), and Bongaerts et al. (2017).

While no broad replication study exists for corporate bonds to the best of our knowledge, replication debates exist on specific papers. Notably, Dickerson et al. (2023) challenge the replicability of the widely used corporate bond factor model of Bai et al. (2019), which was subsequently retracted, and Dickerson, Robotti, and Rossetti (2023) analyze return-based factors. Dickerson et al. (2023) also test for other priced risks, finding that it is "difficult for newly proposed specifications to outperform the simple bond CAPM," echoing earlier findings that the corporate bond market may be well described by the efficient CAPM market (e.g. Fama and French (1993)).

We differ from these papers in several ways. First, while Dickerson et al. (2023) make an important contribution by identifying replication problems with certain factors based on the original methodology, we challenge the general methodologies used in the literature and propose a new framework to come closer to the truth, going beyond detecting replication problems. Second, we consider a broader set of factors, which leads to important differences in economic implications. Indeed, despite finding a low replication rate, we do find a number of significant corporate bond factors, as well as a large number of significant factors based on equity characteristics (where neither of these types of significant factors are considered by Dickerson et al. (2023)) — so we can reject a simple CAPM-like asset pricing model.

In summary, we contribute to the literature by constructing a clean corporate bond data set, proposing a robust factor-construction method, demonstrating a low degree of scientific replication of the corporate bond factor literature, showing which factors remain significant, finding a large number of new significant factors based on equity signals, and thus rejecting a simple CAPM relation for corporate bonds.

1 CRSP for Corporate Bonds: Clean Data Matters

1.1 Cleaning Corporate Bond Returns

The emergence of the TRACE database in 2002, covering almost all transactions in U.S. corporate bonds, has led to a rapidly growing literature on corporate bond factors. The rich dataset, however, has a number of errors, and calculating monthly corporate bond returns is considerably more difficult than calculating monthly equity returns. There are a number of transactions marked as errors in TRACE and these errors are fairly straightforward to delete using standard filters in Dick-Nielsen (2009) and Dick-Nielsen (2014). Nevertheless, after this initial filtering, numerous errors remain, and the literature deals with these errors in different ways.

One approach is to keep all the remaining transactions after the initial filtering. This leaves large errors in the data set as illustrated in Table 1. Panel A shows an obvious price error of 7575.03 for a United Airlines bond, leading to a monthly return of 4585% (we detail the return calculation below). Likewise, Panel B shows that an American Greetings Inc bond has a wrong return of -81% in May 2003 due to an erroneous transaction price of 1. Panel C shows the case of Air Canada, which declared default in April 2003, and, in the early part of the sample period, the physical settlement of credit default swaps is recorded as a transaction with a price at par. Hence, post-default, there are real transactions with prices around 22 to 24 and fictitious transactions at par value 100, leading to an erroneous return in the default month of 172%.

An alternative approach commonly used in the literature is to winsorize returns, for

Panel A: United Airlines 6.01% 01/03/2013 (cusip 909317AY5), return 4585.38% in September 2004				
Date	Time	Price	Is Error	Volume
20/09/2004	10.37	84.25		1,000,000
21/09/2004	11.20	80		15,000
28/09/2004	10.48	7575.03	Error	1,000,000
28/09/2004	13.54	84		896,390
06/10/2004	10.21	84		3,505,000,000
06/10/2004	15.25	84.25		3,131,847

Panel B: American Greetings Inc 6.1% 01/08/2028 (cusip 026375AE5), return -81.13% in May 2003				
Date	Time	Price	Is Error	Volume
23/05/2003	10.42	100.75		50,000
28/05/2003	10.38	101		63,000
28/05/2003	14.27	100		110,000
28/05/2003	15.06	100.5		110,000
28/05/2003	15.21	1	Error	1,510,000
10/06/2003	11.02	101.4375		199,000
10/06/2003	11.03	101.875		199,000
10/06/2003	11.23	102.125		199,000

Panel C: Air Canada 10.25% 15/03/2011 (cusip 008911AJ8), return 171.68% in April 2003				
Date	Time	Price	Is Error	Volume
30/04/2003	11.16	100	Error	9,550,000
30/04/2003	11.17	100	Error	4,550,000
30/04/2003	13.09	23.5		1,500,000
30/04/2003	13.10	23.75		1,500,000
30/04/2003	13.12	24		1,000,000
30/04/2003	13.12	24.25		1,000,000
30/04/2003	14.33	23.75		1,000,000
30/04/2003	14.33	24.125		1,000,000
30/04/2003	15.30	24		500,000
30/04/2003	15.42	100	Error	4,550,000
30/04/2003	16.09	22		1,565,000
30/04/2003	16.09	22.625		1,565,000

Panel D: General Motors Acceptance Corporation 6.8% 15/02/2013 (cusip 37042GV52), return 124.91% in December 2008				
Date	Time	Price	Is Error	Volume
26/11/2008	13.20	21.5		26,000
26/11/2008	13.17	18.5		26,000
30/12/2008	16.15	45		2,340,000
30/12/2008	16.39	41.525		25,000
30/12/2008	17.32	45		2,340,000

Panel E: AMC Entertainment 5.75% 15/06/2025 (cusip 00165AAH1), return 296.77% in January 2021				
Date	Number of trades	Average Price	Is Error	Std. of Price
30/12/2020	21	13.75		0.88
31/12/2020	8	14.54		0.88
04/01/2021	23	15.03		0.17
28/01/2021	135	50.08		1.89
29/01/2021	112	55.87		1.40
01/02/2021	115	55.73		0.82

Table 1: **Examples of Large Returns due to Errors and Non-errors.** This table shows five examples of price sequences in the data set used in the literature (enhanced TRACE after applying the Dick-Nielsen (2014) error filter). Panels A–D report all transactions between the first and last date shown, while Panel E shows average daily prices and standard deviations because there are too many trades to show. Panels A–C contain errors affecting papers without filters, while panels D–E show correct large returns that are filtered away in some papers in the literature.

example at the 99% or 99.5% level and possibly at the 0.5% or 1% level. This approach does not eliminate errors, but reduces their impact. Further, winsorizing also alters correct returns. For example, Panel D of Table 1 shows a bond issued by General Motors Acceptance Corporation with a correct return of 125% in December 2008, but winsorizing at the 99.5% level incorrectly reduces this return to 16.7%. There are 197 General Motors Acceptance Corporation bonds with an average return of 199% in December 2008, so many bond returns are erroneously reduced in this case. As another example shown in Panel E, an AMC Entertainment bond has a correct return of 297% in January 2021 and, again, winsorizing at 99.5% erroneously reduces the return to 16.7%.

We instead solve the problem of data errors by manually checking the data. We start with the enhanced TRACE dataset from July 2002 to December 2021, then delete transactions recorded as errors in TRACE using Dick-Nielsen (2014)'s filter. We exclude floating-rate bonds and merge the TRACE data set with the Mergent FISD dataset that contains information about bond issues and calculate a monthly return for bond i in month t as

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

where $P_{i,t}$ is the transaction price in month t, $AI_{i,t}$ is the accrued interest, and $C_{i,t}$ is the coupon payment, if any, in month t. We compute $P_{i,t}$ as the volume-weighted average price on the last day in month t for which there are any transactions. If there are no transactions in the last five days of month t, $P_{i,t}$ is recorded as missing, resulting in missing values for $r_{i,t}$ and $r_{i,t+1}$.

After calculating monthly returns for our data period, we identify the 5,000 largest returns in absolute value, corresponding to monthly returns where $|r_{i,t}| > 32.6\%$. Every return is investigated independently by two research assistants (RAs), who each assess whether the return observation is 1) an error, 2) not an error, or 3) maybe an error. They are trained to base their decision on a number of factors such as the price pattern and price volatility, odd volumes, prices of other bonds issued by the same firm, the firm's equity price, and internet information about the firm in general. If both RAs agree that the return is an error, the return is recorded as such, and likewise if both RAs agree the return is not an error. If the

RAs disagree, or they both think that the return is maybe an error, one of the authors checks the issue and decides if it is an error or not. We identify 292 errors that we delete, leaving the remaining 4,708 large returns in the dataset. The average (median) absolute return of the errors is 32,000,390% (56%) and thus, these are significant returns that are likely to have a substantial impact on any analysis using the TRACE data set.

In addition to finding a number of errors, the analysis leads to a number of insights on the types of bonds that give rise to errors. We exclude these error-prone bonds as well as bonds that cannot be considered standard corporate bonds. Appendix B provides more details on our methodology.

1.2 Importance of Clean Data: Reproduction

To illustrate the importance of data cleaning choices, we select four well cited corporate bond factors. Each study documents the significance of a specific factor and has a different procedure for cleaning the data.[3] The four factors are momentum (Jostova et al., 2013) (JNPS), value-at-risk (Bai et al., 2019) (BBW), liquidity-beta (Lin et al., 2011) (LWW), and VIX-beta (Chung et al., 2019) (CWW). All these papers use TRACE as the primary data source, but some of the papers also include other data sources.[4] Each paper sorts bonds on a specific signal, creates a long-short portfolio, and the corresponding excess return is denoted the factor return.

For each of these four influential factors, Figure 2 shows the average excess return and its confidence interval, computed in several different ways. First, "Result from paper" (left bar) shows the average factor return and its confidence interval as reported in the original paper.[5]

Second, the orange bar shows our reproduction, that is, the factor return when we use

[3]They differ in how they calculate month-end prices, which bonds to include in their investment universe and whether or not they winsorize/delete returns

[4]Lin et al. (2011) and Chung et al. (2019) use TRACE and NAIC data, Jostova et al. (2013) use the Lehman Brothers Fixed Income Database, Datastream, Bloomberg, TRACE, and NAIC. Bai et al. (2019) rely exclusively on TRACE, although stating that "we relied on a large number of data sources on corporate bonds in addition to TRACE" in Bai et al. (2023) indicates that other sources are used as well. Whenever there are overlapping return observations, all papers use TRACE, so the importance of cleaning data in TRACE is important for all papers.

[5]We use the estimates from Jostova et al. (2013)'s Table 2, Bai et al. (2019)'s Table 2, Lin et al. (2011)'s Table 8 and Chung et al. (2019)'s Table 2.

TRACE data for the same time period, using the same portfolio construction as in the paper, and using the same data cleaning as stated in the paper.[6]

Third, to study the impact of data cleaning, the grey bars in the figure show factor returns computed using the data cleaning used in the *other* papers while keeping everything else the same (i.e., same sample period and portfolio construction). For example, the 'BBW' estimate in the top-left panel shows the momentum return calculated with the data cleaning method of 'BBW' instead of that in the original paper, 'JNPS'.

Finally, the green bars in Figure 2 show factor returns computed using our cleaned data as described as in Section 1.1. Again, we keep everything else as in the original paper to focus specifically on the effect of the data-cleaning method.

As seen in Figure 2, we are only able to reproduce a significant factor return for the momentum factor. While all our reproduction returns have the same signs as the original studies, the results are even more troubling when considering other data-cleaning methods — the factor return often switches sign. For example, while the momentum factor has a statistically significant monthly return of 1.36% using the original filter, the return is -0.19% when using the data cleaning filter in LWW. Looking at our own clean data, we see that none of the factors have significant excess returns.

Overall, Figure 2 shows significant replication problems for the most influential papers in this literature. Further, this figure shows that the choice of data-cleaning procedure is important for the analysis of corporate bond returns. Hence, there is a need for a common framework of clean data, both to ensure that the results are as correct as possible and to reduce the risk of data mining the choices of data-cleaning methods.

Finally, we note that data filters matter even for the most basic statistic, the market return. Indeed, the credit market return is meaningfully different across these methods, and different from the commonly used Bank of America Merrill Lynch corporate bond index return, as seen in the appendix, Figure D.1.

[6]Since we use TRACE data while the other papers use TRACE augmented with other data sources, we expect our replication estimates to be similar but not identical to the original estimate, with the exception of the Value at Risk factor where the estimates should be identical.

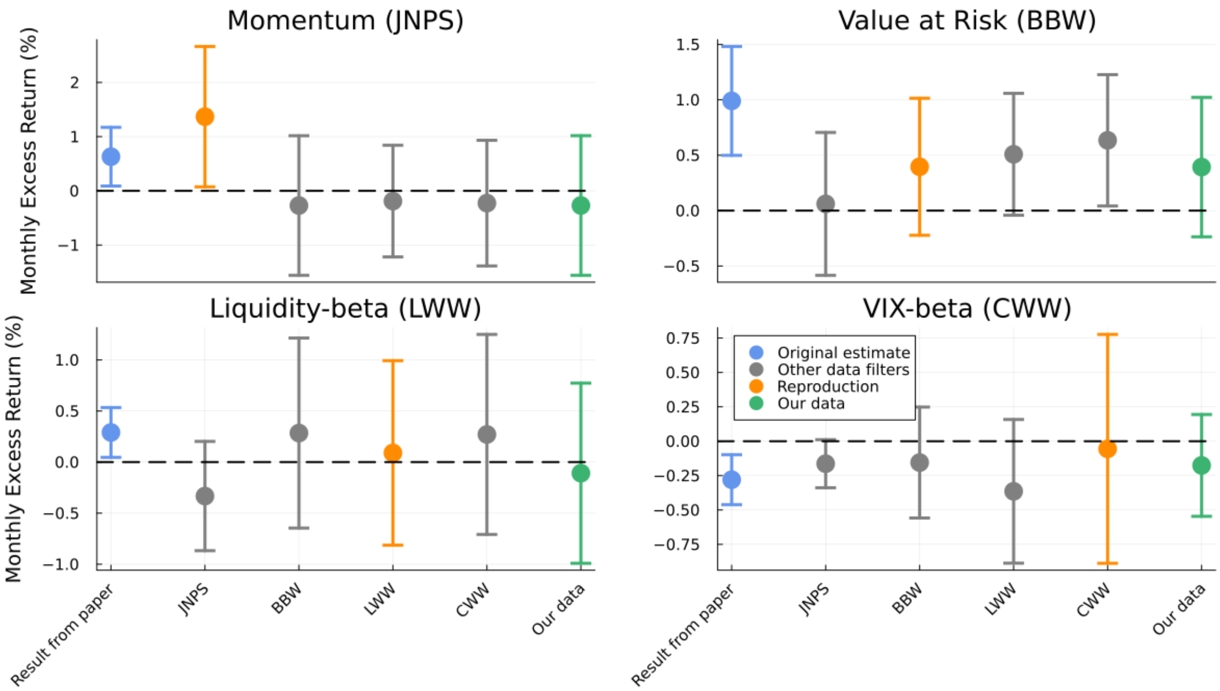

Figure 2: **Replication failure for influential corporate bond factors.** The figure shows average monthly returns for four influential factors in the corporate bond market. For each factor, 'Result from paper' shows the original paper's reported average excess return and its reported confidence interval. For each factor, we replicate the factor construction and sample period from the original paper, but using five different data-cleaning filters, namely those from the four papers ('JNPS', 'BBW', 'LWW', 'CWWW') and our own data cleaning ('Our data'). Here, 'JNPS' refers to Jostova et al. (2013), 'BBW' to Bai et al. (2019), 'LWW' to Lin et al. (2011) and 'CWW' to Chung et al. (2019). The time period is 2002–2011 for Momentum, 2002–2016 for Value at Risk, 2002–2009 for Liquidity-beta and 2002–2016 for VIX-beta. Standard errors are Newey-West adjusted with three lags and the confidence bands are at the 95% level.

2 Scientific Replication: Robust Factor Construction

2.1 Long-Short Factor Construction

In addition to using different data-cleaning methods as discussed in Section 1, the literature also uses different portfolio construction methods. For example, some papers use value-weighting (e.g., Bai et al. (2019)) while others use equal-weighting (e.g., Jostova et al. (2013), Lin et al. (2011), and Chung et al. (2019)). Further, the literature differs in the method of portfolio sorting, with some papers sorting into 10 portfolios while others use fewer portfolios, some papers double-sort while others single-sort, and some use monthly rebalancing while others rebalance less frequently.

We want a robust and consistent method for constructing all long-short factor returns, in addition to our clean data. Our factor returns are constructed as follows. Each month, we first sort all bonds into three ratings categories:

- Investment grade plus, "IG+," i.e., the highest rated bonds with ratings from AAA to A3 in Moody's or AAA to A- in Standard&Poor's (if these rating agencies disagree, we take the average numeric rating, from 1 to 22, and round up, i.e. using the more conservative rating if they only disagree by one number).

- Investment grade minus, "IG-," i.e., bonds with ratings from Baa1 to Baa3 in Moody's or BBB+ to BBB- in Standard&Poor's (again using the average numeric rating).

- Speculative grade, "SG," i.e., all bonds rated Ba1 or lower in Moody's or BB+ or lower in Standard&Poor's.

Within each of these three ratings categories c, we sort all bonds into three groups based on the signal of interest i —– for example, yield-to-maturity (YTM), momentum, or bond age (i.e., using a dependent sort) — and compute value-weighted returns, $r_t^{i,c}$, for each signal-by-rating group. Using these returns, we compute long-short factor returns within each rating category. For example, for factor i in IG+, we buy bonds with a high signal i while shorting those with low signals within IG+:

$$f_t^{i,IG+} = r_t^{High\text{-}i,IG+} - r_t^{Low\text{-}i,IG+} \tag{2}$$

and similarly for $f_t^{i,IG\text{-}}$ and $f_t^{i,SG}$. Finally, our overall factor return for any signal i is the average of the three factor returns in each rating category:

$$f_t^{i} = \frac{1}{3}\left(f_t^{i,IG+} + f_t^{i,IG\text{-}} + f_t^{i,SG}\right) \tag{3}$$

This methodology is motivated by the following considerations. First, we use value-weighted returns since this makes the factors more implementable than equal-weighting and reduces issues with missing values (see Table B.2 in Appendix B). Second, we use tertile portfolios, i.e., sort into three groups, since this uses two-thirds of the data, rather than the more extreme decile portfolios often used in the literature that only uses 20% of the data. Third, we double-sort based on each signal and the three broad credit-rating groups to ensure apples-to-apples comparisons while simultaneously not slicing the data too finely (similar to the Fama-French construction for equity factors). Lastly, we use monthly rebalancing to condition on the most up-to-date information.

Figure 3 shows an example of how portfolio construction can matter. The figure replicates four factors from the literature using a common sample (our clean data over the TRACE period), only varying the portfolio construction. The figure shows the excess return of each factor when the portfolio construction is, respectively, our method or three different methods from the literature. As seen in the figure, the portfolio construction method has a meaningful effect on both the point estimate and its standard error. Interestingly, our method has the smallest standard errors, thus providing the most precise estimates. Our method's smaller standard errors may be due to a reduction of noise via value weighting, apples-to-apples comparisons within the rating groups, and the broad tercile sorts.

In summary, Figure 3 shows that portfolio construction matters even with common data, just like Figure 2 shows that data matters even with common portfolio construction. We focus from now on our common data and common methodology.

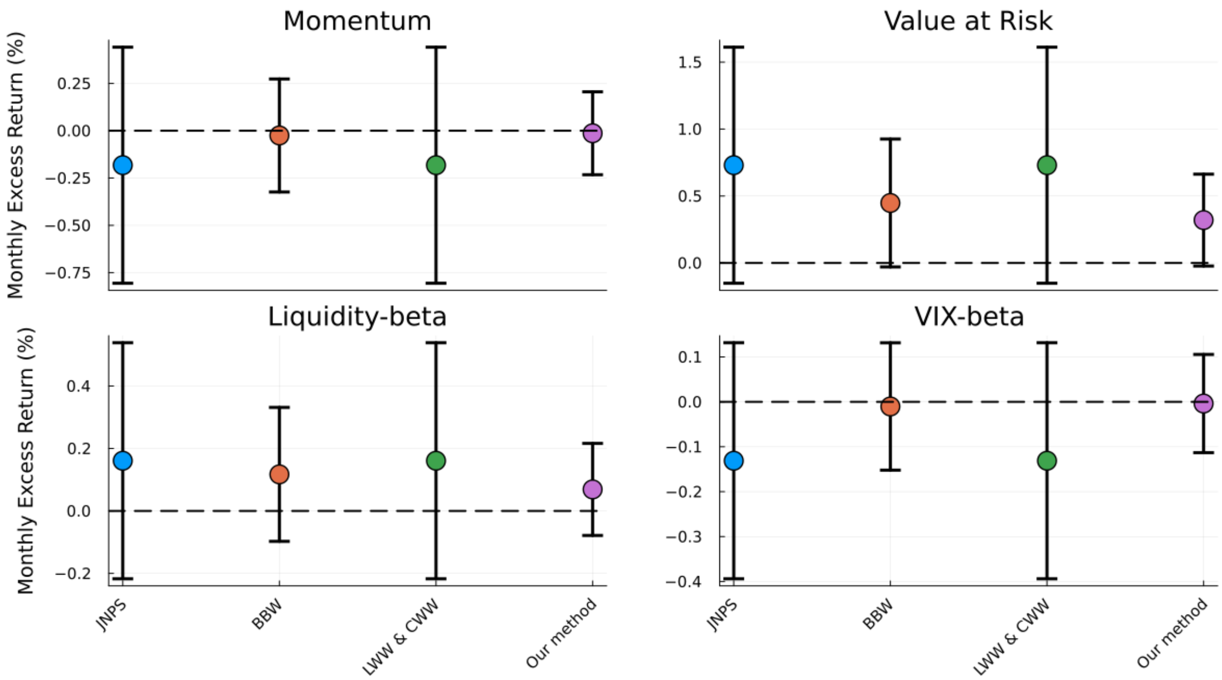

Figure 3: **Factor construction impact on influential corporate bond factors.** The figure shows average monthly excess returns for four influential factors in the corporate bond market. For each factor, we use our data and TRACE sample period to replicate each signal, but then use the factor construction used in the four papers as well as our own method, where 'JNPS' refers to Jostova et al. (2013), 'BBW' to Bai et al. (2019), 'LWW' Lin et al. (2011), and 'CWW' to Chung et al. (2019). Standard errors are Newey-West adjusted with three lags and the confidence bands are at the 95% level.

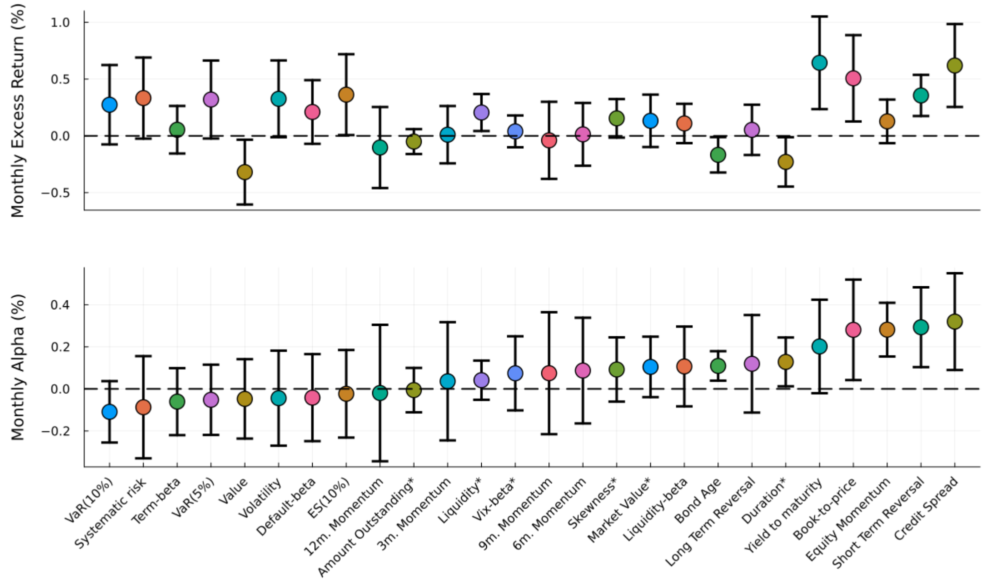

Figure 4: **Bond factor replication.** The top panel in the figure reports corporate bond factors' monthly excess returns and their confidence intervals. The bottom panel reports the corresponding alphas and their confidence intervals, where alpha is the intercept in a regression of monthly excess returns on the corporate bond market factor and the Fama and French (1993) TERM factor. The sample period is July 2002–December 2021. Factor returns are computed each month from value-weighted corporate bond returns for portfolios that are double-sorted into three rating groups and three characteristic groups. Each factor is long the high characteristic portfolio and short the low characteristic portfolio, except factors indicated by "*" where the factor is long the low characteristic. Standard errors are Newey-West adjusted.

2.2 Replication Results

Figure 4 reports our scientific replication of all the corporate bond factors. The top panel shows each factor's excess return and its confidence interval. Similarly, the bottom panel shows each factor's alpha and its confidence interval, where the alpha is the intercept from the following regression of monthly factor excess returns on our corporate bond market factor, $CMKT$, and the $TERM$ factor:

$$f_t^i = \alpha^i + \beta_{CMKT}^i CMKT_t + \beta_{TERM}^i TERM_t + \varepsilon_t \tag{4}$$

This alpha captures a factor's excess return when controlling for the compensation for credit risk that can be achieved simply by buying corporate bonds in general ($CMKT$) and the general compensation for duration risk that can be achieved via long-term government bonds

($TERM$).[7] Hence, we focus on the significance of alphas (rather than excess returns).

As seen in Figure 4, most of the factors considered in the literature are insignificant, that is, fail to replicate. The factors that do replicate are bond age, short duration, book-to-price, equity momentum, short-term reversal, and credit spread.

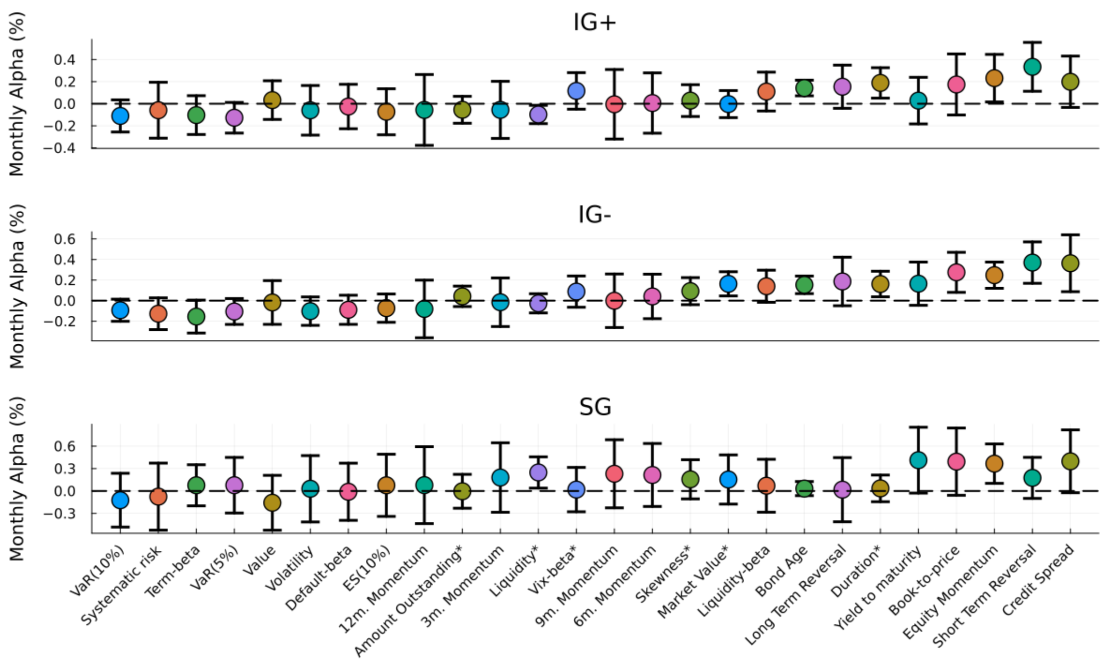

Figure 5: **Bond factor replication: Breakdown on rating category.** The figure reports corporate bond factors' monthly alphas and their confidence intervals in three rating categories: IG+ (top panel), IG- (middle), and SG (bottom). Alpha is the intercept in a regression of monthly excess returns on the corporate bond market factor and the Fama and French (1993) TERM factor, and the sample period is July 2002–December 2021. Factor returns are computed each month from value-weighted corporate bond returns for portfolios that are sorted into three characteristic groups. Each factor is long the high characteristic portfolio and short the low characteristic portfolio, except factors indicated by "*" where the factor is long the low characteristic. Standard errors are Newey-West adjusted.

Similarly, Figure 5 reports the alphas of factors within each rating category, showing how factors perform among bonds of different credit risk. For ease of comparison, the factors are listed in the same order in Figures 4 and 5 (in both cases sorted on the magnitude of the alpha depicted in Figure 4). We see that the factors that were insignificant in Figure 4 tend to be

[7]Specifically, we calculate $CMKT$ as the value-weighted average excess return of all corporate bonds in our sample. Further, $TERM$ is the average return on Treasury bonds with maturities greater than 120 months from CRSP's Fama Maturity Portfolios, in excess of the T-Bill rate from Ken French's website, similar to the $TERM$ factor in Fama and French (1993).

insignificant in all rating groups. The results for the significant factors tend to be consistent across rating groups, but the significance is lower in some rating groups, which could simply be due to more noisy returns when the amount of data is smaller. Significance is particularly strong among IG- bonds, where several other factors become marginally significant (long-term reversal, market value, and yield to maturity).

2.3 Robustness

As a robustness analysis, Figure 6 shows the factor alphas with three alternative methodologies. In particular, we consider lagged signals, a liquid universe, and pure credit returns.

Panel A of Figure 6 reports the factor alphas when the signals have been lagged a month. As a way of background for this analysis, note that the results in Section 2.2 compute each signal at the end of month t for predicting returns in the month from t to $t + 1$, so the same price may be used to calculate the signal and the denominator in the return, as seen in equation (1). This convention used in the literature is not entirely realistic from a trading perspective, but can be justified from an academic perspective if we are simply interested in understanding drivers of returns. In contrast, when signals are lagged a month as in Panel A, the returns are easily implementable, but arguably overly conservative from a trading perspective.

In any event, we see that the performance of several of the significant factors is reduced, both in terms of the alpha's point estimate and its significance. The most affected factor is short-term reversal, where the alpha becomes negative and insignificant, perhaps because this factor is largely driven by an unimplementable bid-ask bounce when the signal is not lagged. Only a couple of factors remain significant with lagged signals (bond age, equity momentum) and one is marginally significant (duration).

Panel B shows the results with a liquid universe of bonds from large firms that frequently trade. Each month, we form portfolios only using the bonds from the 50% largest issuers based on total amount outstanding across all corporate bonds (in that month) and further condition on these bonds having non-missing returns over the prior six months. This liquid sample of corporate bonds should be more realistic to implement as a portfolio strategy. The results to a large extent mirror those from the broader bond universe in Figure 4, suggesting

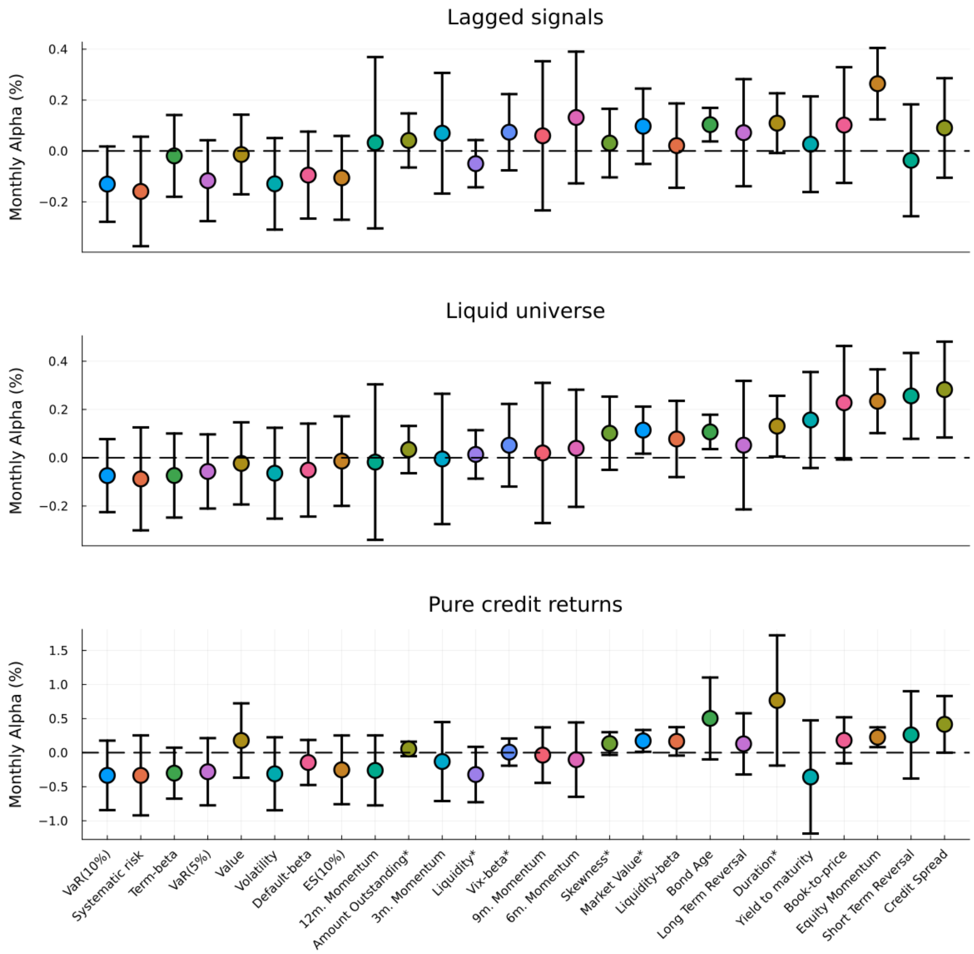

Figure 6: **Bond factor replication: Robustness.** The figure reports corporate bond factors' monthly alphas and their confidence intervals, where alpha is the intercept in a regression of monthly excess returns on the corporate bond market factor and the Fama and French (1993) TERM factor. In the top panel, signals are lagged by a month, in the middle panel we only include bonds from large firms that frequently trade and in the bottom panel we hedge each bonds' promised payments using a matched treasury portfolio. The sample period is July 2002–December 2021. Factor returns are computed each month from value-weighted corporate bond returns for portfolios that are double-sorted into three rating groups and three characteristic groups. Each factor is long the high characteristic portfolio and short the low characteristic portfolio, except factors indicated by "*" where the factor is long the low characteristic. Standard errors are Newey-West adjusted.

that the alpha of the significant factors is not primarily driven by stale prices and bonds issued by small issuers.

As another robustness check, we compute corporate bond returns using a more sophisticated interest-rate hedging method in the spirit of van Binsbergen and Schwert (2022) and Andreani et al. (2023). In particular, for each bond, we hedge each of its promised payments using the corresponding zero-bond Treasury rate. We then use these hedged bond returns as test assets in the same way as in the baseline specification (i.e., using exactly the same signals as before). Panel C of Figure 6 shows the resulting factor alphas, again showing a modest replication rate.

2.4 Out-of-Sample Evidence: Factor Returns Before Trace

The TRACE data has enabled a lot of research on corporate bonds, but it only starts in 2002. Hence, the factor returns based on this data have a much shorter sample, currently around 20 years, than the typical studies of equity factors where the US data begins in 1926 and data in most other developed countries begins in 1986 (Jensen et al., 2022).

Considering a longer sample is interesting in its own right, and, further, the data before TRACE can be seen as out-of-sample evidence for the factors based on papers that only considered TRACE data. Figure 7 shows the factor performance over the early time period 1985-2002 using a combination of the so-called Warga dataset (1973–1998) and the Merrill Lynch/ICE dataset for (1997–2002), which are based on quotes, not trades. Details about these datasets are in Appendix C. These data are available only monthly, which means that certain characteristics cannot be computed because they require daily data (liquidity, liquidity-beta, systematic risk, and Vix-beta).

As seen in Figure 7, there are six significantly positive factors at the bond level in the early sample, namely amount outstanding, market value, bond age, long-term reversal, duration, and equity momentum. Most of the other factors are insignificant, but a few are even significant with the opposite sign of that proposed in the literature.

Figure 7 also reports the performance over the merged sample from 1985 until 2021. This sample also has six significant factors. These are the same as in the early sample, except that credit spread becomes significant and amount outstanding becomes insignificant. Other

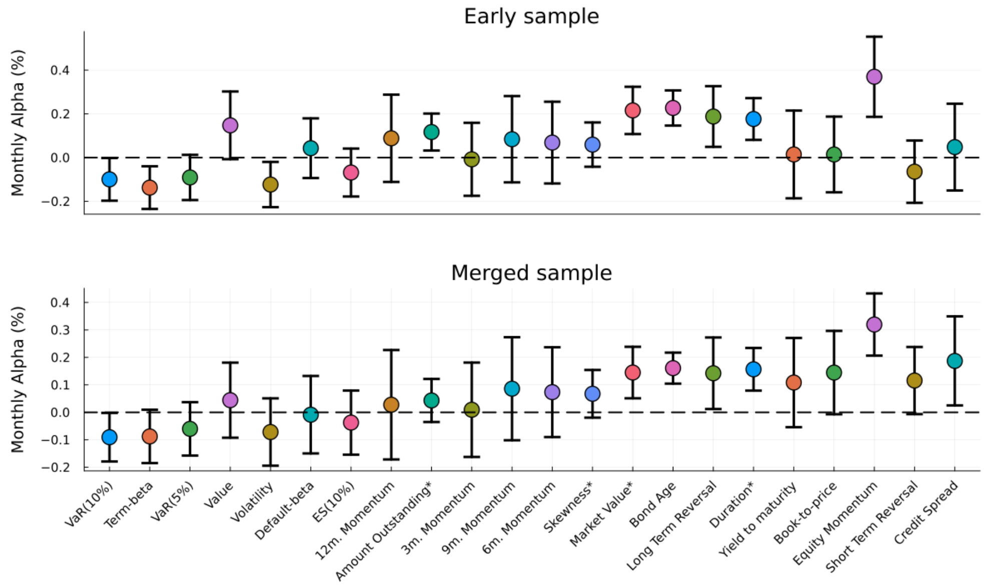

Figure 7: **Out-of-sample evidence: Factor alphas over the early sample and merged samples.** The figure reports monthly alphas and their confidence intervals for bond-level factors over the early sample 1985-2002 (top panel) and merged sample 1985-2021 (bottom panel). The early data comes from the Warga- and Merill Lynch/ICE dataset, both of which is based on quotes, and the recent data from 2002-2021 comes from TRACE, which is based on trades.

than these six significantly positive factors, all of the other factors are insignificant, except that VaR(10%) is marginally significantly negative.

2.5 Multiple-Testing Adjustments

Given that we are testing the significance of each of many factors, accounting for multiple testing (MT) problems is prudent. For example, even if none of the factors were truly significant, 5% of them would appear to be significant just by chance when using the standard tests with a cut-off of 1.96 for t-statistics.

To account for such MT problems, we use the most standard frequentist MT adjustment from Benjamini and Hochberg (1995). As mentioned in the introduction, such MT adjustments depend on the collection of all the tests tried by researchers, both those that are published and those that are not. While knowing all tried tests is never possible, we use a comprehensive set based on all corporate bond factors at the bond level (as in Section 2.2) and at the firm level (defined in Section 3.2), as well as all factors based on equity signals (defined in Section 3.3) resulting in a total of 197 factors. Hence, our equity-based signals are both interesting in their own right and a useful device for MT testing correction as they serve as a proxy for other factors that researchers might have tried. Using Benjamini and Hochberg (1995) for this set of tests implies that a factor is considered significant if its alpha has an absolute t-statistic greater than 2.43.

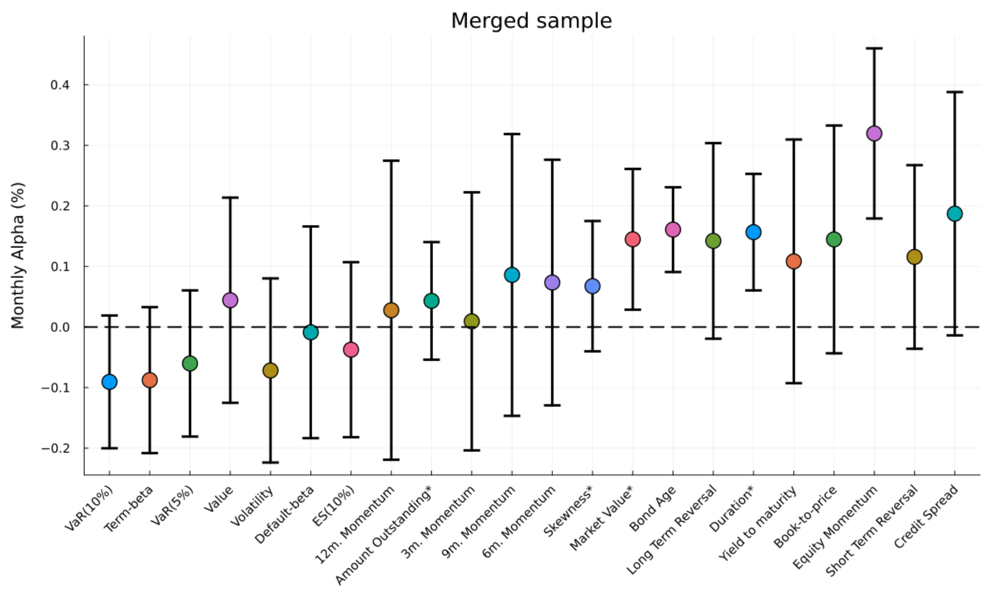

Figure 8: **Significance with Multiple-Testing Adjustment: Merged Sample.** The figure reports monthly alphas and their confidence intervals, adjusted for multiple-testing effects using Benjamini and Hochberg (1995). The results are for bond-level factors over the merged sample, 1985-2021.

Figure 8 reports the confidence intervals using the MT adjustment for the bond-level factors over the merged sample. Specifically, while the standard confidence interval is the alpha plus and minus 1.96 times the standard error of the alpha, Figure 8 plots instead each estimated alpha plus and minus 2.43 times the standard error.

Each factor is only considered significant if the entire bar is above zero (as before), and now the confidence interval is wider due to the MT correction. The factors that are significant with MT correction are bond age, market value, duration, and equity momentum.

3 New Firm-Level Returns and Factors

3.1 Constructing Firm-Level Returns and Signals

We are also interested in testing which factors are significant when using firm-level bond returns. Firm-level bond returns have the advantages that (i) each firm has a single time series of representative bond returns; (ii) firms with many bonds are not over-represented when setting thresholds for tertile portfolios; (iii) this method allows us to separate whether factors are driven by firm-level or bond-level characteristics; and (iv) the number of missing return observations is reduced (as seen Table B.2 in Appendix B).

We calculate the overall corporate bond return for firm k in month $t+1$ as

$$r_{t+1}^{k} = \sum_{i \in \mathcal{R}} w_t^{k,i} r_{t+1}^{k,i} \tag{5}$$

where $\mathcal{R}$ is the set of the firm's bonds with non-missing returns during this month, $r_{t+1}^{k,i}$ is the return of the firm's i'th bond, and $w_t^{k,i} = mv_t^{k,i} / \sum_{j \in \mathcal{R}} mv_t^{k,j}$ is the corresponding fraction of the market value (mv) at the end of the previous month t.

If a firm has a bond z with a missing return in month $t+1$, then we impute this missing return from the observed average bond return, r_{t+1}^{k}, using a duration adjustment, $r_{t+1}^{k,z} = r_{t+1}^{k} \frac{D_t^z}{D_t^{\mathcal{R}}}$, where D_t^z is the duration of the bond with missing return and $D_t^{\mathcal{R}} = \sum_{i \in \mathcal{R}} D_t^{k,i} \frac{mv_t^{k,i}}{\sum_{j \in \mathcal{R}} mv_t^{j,k}}$ is the overall duration of the bonds with observed returns.

When aggregating all bond returns to the firm level, including both those with observed returns and those with imputed returns, we get the following firm-level excess return adjusted for missing bonds:

$$\tilde{r}_{t+1}^{k} = r_{t+1}^{k} \frac{D_t^{\mathcal{A}}}{D_t^{\mathcal{R}}} \tag{6}$$

where $D_t^{\mathcal{A}}$ is the average duration across the set of all bonds, $\mathcal{A}$, with observed returns or imputed returns.[8]

We compute the firm-level characteristics as follows. For signals that are functions of re-

[8] To see that adjusting each bond return implies (6) at the overall firm level, let $\mathcal{N} = \mathcal{A} \setminus \mathcal{R}$ be the bonds

turns (e.g., momentum and default-beta), we use the firm-level returns to compute the signal. For other signals (e.g., credit spread), we compute the firm-level signal as the value-weighted average of the bond-level signals, i.e., $s_t^k = \sum_i s_t^{i,k} w_t^{i,k}$ as in Equation (5), except that we naturally take sums instead of value-weighted averages in the special cases of "amount outstanding" and "market value."

3.2 Firm-Level Factors based on Existing Corporate Bond Signals

We construct factors using the same characteristics and constructed in the same way as described in Section 2.1, except that now we use returns and signals constructed at the firm-level, instead of at the bond-level, as described in Section 3.1.

The returns of these firm-level factors are shown in Figure 9. We see that bond age and equity momentum remain significant as in Figure 4 while some factors lose their significance at the firm level, namely duration, book-to-price, short-term reversal, and credit spread.

Interestingly, the size-related factors based on low market value and low amount outstanding become significant at the firm level, while these were insignificant at the bond level in Figure 4. To understand this finding, we note that large firms may have many "normal sized" bond issues, so large firms are better identified by the total value (or amount outstanding) of all their bonds (i.e., size at the firm level) than their size at the bond level. So, our finding means that bonds issued by firms with a smaller overall footprint in the corporate bond market have larger returns, but the size of the specific bond issue does not appear important. An explanation could be that bond liquidity increases in the overall size of the whole collection of the firm's bonds, and investors require compensation for holding illiquid bonds or bonds with liquidity risk. Another liquidity-related factor is bond age, which is

with missing return (but observed duration and market value) and note that:

$$
\begin{aligned}
\tilde{r}_{t+1}^k &= \sum_{i\in\mathcal{A}} r_{t+1}^{k,i} \frac{mv_t^{k,i}}{\sum_{j\in\mathcal{A}} mv_t^{k,j}} = \sum_{i\in\mathcal{R}} r_{t+1}^{k,i} \frac{mv_t^{k,i}}{\sum_{j\in\mathcal{A}} mv_t^{k,j}} + \sum_{i\in\mathcal{N}} r_{t+1}^{k,i} \frac{mv_t^{k,i}}{\sum_{j\in\mathcal{A}} mv_t^{k,j}} \\
&= r_{t+1}^k \frac{\sum_{j\in\mathcal{R}} mv_t^{k,j}}{\sum_{j\in\mathcal{A}} mv_t^{k,j}} + r_{t+1}^k \sum_{i\in\mathcal{N}} \frac{D_t^i}{D_t^{\mathcal{R}}} \frac{mv_t^{k,i}}{\sum_{j\in\mathcal{A}} mv_t^{k,j}} = r_{t+1}^k \frac{1}{D_t^{\mathcal{R}}} \left(D_t^{\mathcal{R}} \frac{\sum_{j\in\mathcal{R}} mv_t^{k,j}}{\sum_{j\in\mathcal{A}} mv_t^{k,j}} + D_t^{\mathcal{N}} \frac{\sum_{i\in\mathcal{N}} mv_t^{k,i}}{\sum_{j\in\mathcal{A}} mv_t^{k,j}} \right) \\
&= r_{t+1}^k \frac{D_t^{\mathcal{A}}}{D_t^{\mathcal{R}}}
\end{aligned}
$$

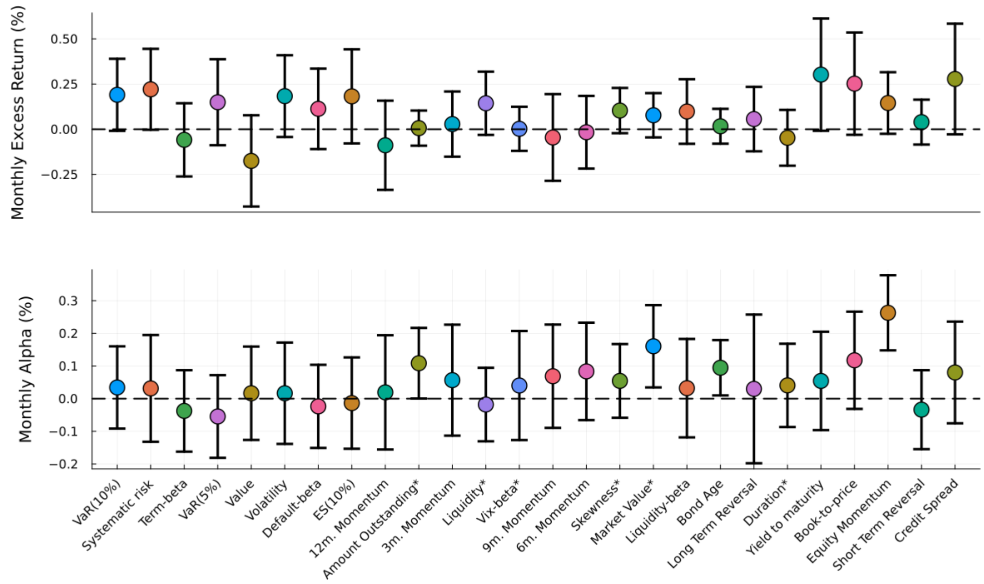

Figure 9: **Factor performance at the firm level.** Using factors constructed based on overall firm-level corporate bond returns and the corresponding firm-level characteristics, the top panel reports monthly excess returns and their confidence intervals. The bottom panel reports the corresponding alphas and their confidence intervals, where alpha is the intercept in a regression of monthly excess returns on the corporate bond market factor and the Fama and French (1993) TERM factor. The sample period is July 2002–December 2021. Factor returns are computed each month from value-weighted portfolios that are double-sorted into three rating groups and three characteristic groups. Each factor is long the high characteristic portfolio and short the low characteristic portfolio, except factors indicated by "*" where the factor is long the low characteristic. Standard errors are Newey-West adjusted.

significant both at the bond and firm level.

3.3 New Corporate Bond Factors: Signals from Equities

Following Fama and French (1993) and Chordia et al. (2017), it is interesting to consider how signals used to predict equity returns perform in terms of predicting corporate bond returns. To analyze this question in a broad way, we consider all the 153 equity signals from Jensen et al. (2022). The signals are grouped into 13 clusters based on their pairwise correlation of CAPM-residual returns. The signals are signed so as to predict equity returns with a positive sign, but we note that only some of these factors are statistically significant predictors of equity returns as explained in Jensen et al. (2022).

The bottom panel of Figure 10 reports the alphas of all these 153 corporate bond factors over the merged sample period, 1985–2021. Further, the top panel of Figure 10 shows the performance of the average factor returns within each of the 13 clusters.[9]

We compute the confidence interval of each of the alphas of factors and cluster returns using a multiple-testing adjustment as explained in Section 2.5. Specifically, each confidence interval is the estimated alpha plus and minus 2.43 times the standard error (i.e., a wider confidence interval than using the standard multiplier of 1.96).

While most of these factors are new in a corporate bond context, Chordia et al. (2017) already considered 10 of these. Our results are broadly consistent with those of Chordia et al. (2017), although this paper does not report alphas of all factors. For example, Chordia et al. (2017) find that 6m equity momentum is a positive predictor of bond returns, consistent with the other literature (e.g., Gebhardt et al. (2005b), Jostova et al. (2013)) and our replication. They also find that short-term equity reversal is a negative predictor of bond returns, meaning that equity returns over the past month predict bond returns positively, consistent with our findings in Figure 10.

Figure 10 finds a surprisingly large number of significant factors. In particular, we find 36 significantly positive factors and 17 significantly negative. In other words, while the literature on corporate bond factors includes a large number of factors that we cannot replicate, it does not include a large number of equity-based factors that we find to be significant.

[9] Figure D.4 in the appendix shows the corresponding results over the TRACE sample.

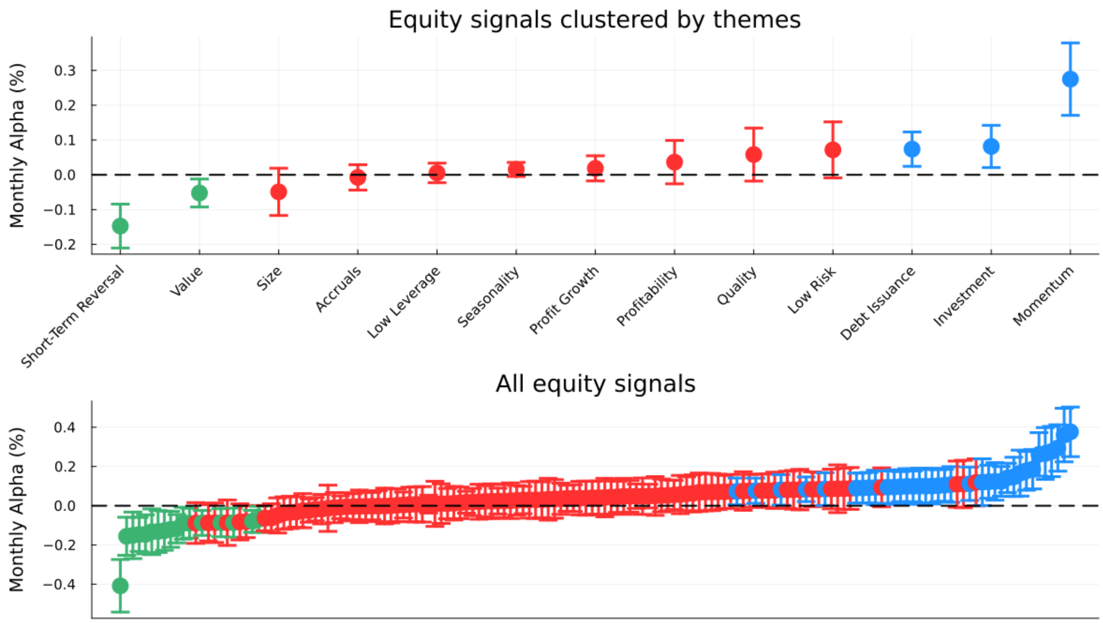

Figure 10: **Corporate bond factors based on equity signals: Merged sample 1985-2021 with BH adjustment.** The top panel reports the monthly alphas of the average factor returns within each cluster and their confidence intervals adjusted for multiple-testing (MT) effects using Benjamini and Hochberg (1995). Similarly, the bottom panel reports monthly alphas and their MT-adjusted confidence intervals for each of the individual corporate bond factors based on the 153 equity signals from Jensen et al. (2022). As seen in the bottom panel, 36 factors are significantly positive (blue dots), 17 are significantly negative (green dots), and 100 are insignificant (red dots).

As a robustness check, we also compute the number of significant alphas using an alternative multiple-testing adjustment. In particular, Figure D.5 in the appendix reports the results using the empirical-Bayes methodology of Jensen et al. (2022). Using this methodology implies that 34 factors are significantly positive and 24 are significantly negative, similar to our results with the MT-adjustment based on Benjamini and Hochberg (1995).

Looking at the cluster results in Figure 10, we see that corporate bonds are positively predicted by equity momentum, including short-term equity momentum (the negative of short-term reversal), investment, and debt issuance. Further, while a cheap stock value is a positive predictor of equity returns, it turns out to be a negative predictor of corporate bond returns, which may be related to risk-shifting or financial distress of these firms.

4 Conclusion

Many corporate bond factors cannot be reproduced even when attempting to use the methodology of the corresponding paper. More broadly, even factors that can be reproduced should be questioned, since the corporate bond literature is based on data full of errors.

To come closer to the truth about the drivers of corporate bond returns, we put together a database of cleaned corporate bond returns based on manual checks of all the largest returns, among other corrections. Further, we propose a common methodology for factor construction, which can be applied both at the level of individual bonds and at the level of representative firm-level corporate bond returns.

Using this relatively clean data and robust methodology, we show that the majority of corporate bond factors from the literature fail to replicate, but a minority of factors remain significant. Further, analyzing corporate bond factors based on equity signals, we find a number of significant new factors.

These findings challenge most of the rapidly growing literature on corporate bond factors, and simultaneous challenge the opposite view that the CAPM largely works for corporate bonds (Fama and French (1993), Dickerson et al. (2023)). We hope that our clean corporate bond returns, clean firm-level returns, and code will enable future research that brings the profession even closer to the truth about the corporate bond market.

References

Acharya, V. V., Y. Amihud, and S. T. Bharath (2013). Liquidity risk of corporate bond returns: conditional approach. *Journal of financial economics 110*(2), 358–386.

Andreani, M., D. Palhares, and S. Richardson (2023). Computing corporate bond returns: a word (or two) of caution. *Review of Accounting Studies*, 1–20.

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

Bai, J., T. G. Bali, and Q. Wen (2021, 12). Is there a risk-return tradeoff in the corporate bond market? time-series and cross-sectional evidence. *Journal of Financial Economics 142*, 1017–1037.

Bai, J., T. G. Bali, and Q. Wen (2023, July). Replication of bbw factors with wrds data. *Available at SSRN 4476612*.

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

Bartram, S. M., M. Grinblatt, and Y. Nozawa (2019, 12). Book-to-market, mispricing, and the cross-section of corporate bond returns. *SSRN Electronic Journal*.

Benjamini, Y. and Y. Hochberg (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.

Bettis, R. A. (2012). The search for asterisks: Compromised statistical tests and flawed theories. *Strategic Management Journal 33*(1), 108–113.

Bongaerts, D., F. De Jong, and J. Driessen (2017). An asset pricing approach to liquidity effects in corporate bond markets. *The Review of Financial Studies 30*(4), 1229–1269.

Chen, A. Y. and T. Zimmermann (2020). Publication bias and the cross-section of stock returns. *The Review of Asset Pricing Studies 10*(2), 249–289.

Chordia, T., A. Goyal, Y. Nozawa, A. Subrahmanyam, and Q. Tong (2017). Are capital market anomalies common to equity and corporate bond markets? an empirical investigation. *Journal of Financial and Quantitative Analysis 52*(4), 1301–1342.

Chung, K., J. Wang, and C. Wu (2019). Volatility and the cross-section of corporate bond returns. *Journal of Financial Economics 133*(2), 397–417.

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

Dick-Nielsen, J. (2014). How to Clean Enhanced TRACE Data. *Unpublished Manuscript*.

Dickerson, A., P. Mueller, and C. Robotti (2023). Priced risk in corporate bonds. *Journal of Financial Economics 150*(2).

Dickerson, A., C. Robotti, and G. Rossetti (2023). Return-based anomalies in corporate bonds: Are they there? *Available at SSRN 4575879*.

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

Gebhardt, W. R., S. Hvidkjaer, and B. Swaminathan (2005a, 1). The cross-section of expected corporate bond returns: Betas or characteristics? *Journal of Financial Economics 75*, 85–114.

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

Greenwood, R. and S. Hanson (2013). Issuer Quality and Corporate Bond Returns. *Review of Financial Studies 26*(6), 1483–1525.

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *The Review of Financial Studies 29*(1), 5–68.

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

Houweling, P., A. Mentink, and T. A. Vorst (2005, 11). Comparing possible proxies of corporate bond liquidity. *SSRN Electronic Journal*.

Houweling, P. and J. V. Zundert (2018). Factor investing in the corporate bond market. *https://doi.org/10.2469/faj.v73.n2.1 73*, 100–115.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine 2*(8), e124.

Israel, R., D. Palhares, and S. A. Richardson (2018). Common factors in corporate bond returns. *Journal of Investment Management 16*(2), 17–46.

Jensen, T. I., B. T. Kelly, and L. H. Pedersen (2022). Is there a replication crisis in finance? *Journal of Finance, forthcoming*.

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

Kelly, B., D. Palhares, and S. Pruitt (2023). Modeling corporate bond returns. *The Journal of Finance 78*(4), 1967–2008.

Lin, H., J. Wang, and C. Wu (2011). Liquidity risk and expected corporate bond returns. *Journal of Financial Economics 99*, 628—-650.

Maniadis, Z., F. Tufano, and J. A. List (2017). To replicate or not to replicate? exploring reproducibility in economics through the lens of a model and a pilot study. *The Economic Journal 127*, F209–F235.

Nosek, B. A., J. R. Spies, and M. Motyl (2012). Scientific utopia: Ii. restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science 7*(6), 615–631.

van Binsbergen, J. H. and M. Schwert (2022). Duration-based valuation of corporate bonds. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper*.

Van Mencxel, K. (2022). The cross-section of corporate bond returns: Evidence from an elusive past. *Job Market Paper*.

Appendix

A Details on Factors

Name	Description	Cite	Direction
VaR(5%)	-1 * 2nd lowest observation in a 36 month rolling window	Bai et al. (2019)	1
VaR(10%)	-1 * 4th lowest observation in a 36 month rolling window	Bai et al. (2019)	1
ES(10%)	-1 * mean of 4 lowest obs. in 36 month rolling window	Bai et al. (2019)	1
Credit Rating	Numerical credit rating from 1-21	Bai et al. (2019)	1
Liquidity*	-1 * covariance of daily returns within a month	Bai et al. (2019)	1
Liquidity-beta	The regression coefficient on liquidity in a rolling regression of bond excess return on stock market excess return, SMB, HML, DEF, TERM and Amihud Illiquidity factors	Lin et al. (2011)	1
Short Term Reversal	Return from t-1 to t, sign flipped	Bai et al. (2019)	1
Volatility	36 month vol	Bai et al. (2019)	1
Skewness*	36 month skew	Bai et al. (2019)	-1
Kurtosis	36 month kurt	Bai et al. (2019)	1
Book-to-price	Bond book value / market price	Bartram et al. (2019)	1
Systematic risk	Total var minus var on residual from factor regression	Bai et al. (2021)	1
3m. Momentum	3 month momentum	Jostova et al. (2013)	1
6m. Momentum	6 month momentum	Jostova et al. (2013)	1
9m. Momentum	9 month momentum	Jostova et al. (2013)	1
12m. Momentum	12 month momentum	Jostova et al. (2013)	1
Duration*	Modified Macaulay duration	Israel et al. (2018)	-1
Long Term Reversal	Loser minus winners from t-48 to t-13	Bali et al. (2021)	1
Vix-beta*	The regression coefficient on VIX innovations in a rolling regression of bond excess return on stock market excess return, SMB, HML, DEF, TERM, Amihud Illiquidity and VIX factors.	Chung et al. (2019)	-1
Default-beta	The regression coefficient on DEF in a rolling regression of bonds excess return on the DEF factor.	Gebhardt et al. (2005a)	1
Term-beta	The regression coefficient on TERM in a rolling regression of bonds excess return on the TERM factor.	Gebhardt et al. (2005a)	1
Yield to maturity	Promised yield to maturity	Gebhardt et al. (2005a)	1
Credit Spread	Yield to maturity minus the yield on a cash-flow matched portfolio of treasuries	Israel et al. (2018)	1
Bond Age	Fraction of bond life expired	Houweling et al. (2005) & Van Mencxel (2022)	1
Amount Outstanding*	Bond notional * number of bonds outstanding	Houweling et al. (2005)	-1
Market Value*	Price * Amount Outstanding	Houweling and Zundert (2018)	-1
Equity Momentum	6 month equity momentum	Gebhardt et al. (2005b) & Israel et al. (2018)	1
Value	The residual from a cross-sectional regression of the log of credit spread onto the log of duration, rating and bond excess return volatility (12 month)	Israel et al. (2018)	1

Table A.1: **Factor Details.** This table shows the corporate bond factors used in our study, the reference for each factor, the sign of the effect in the original paper, and an indicator of the significance in the original study.

B Cleaning TRACE Data

To clean the TRACE data, we apply a number of filters described in Table B.1. We begin by calculating monthly prices using Enhanced TRACE and merge them with bond characteristics from Mergent FISD using bond CUSIP. We exclude fixed rate bonds with inconsistent interest payment dates recorded in FISD and with a coupon frequency above 12. We then calculate returns and exclude returns from variable rate bonds and bonds marked with the convertible flag in FISD. Next, we delete the manually detected return errors. Based on the error analysis we further exclude convertible bonds identified as bond type CCOV in Mergent FISD and bonds that do not have a principal value of $1,000. We exclude bonds that are trust-preferred capital securities (TPCS), US corporate PIK bonds (CPIK), US Corporate Pass Thru Trust (CPAS), equity-linked and non-US non-corporate bonds. We define equity linked bonds, as bonds whose field 'issue name' contains any of the strings 'EQUITY-LINKED', 'EQUITY LINKED' 'INDEX-LINKED'. Non-US non-corporate bonds consist of bonds whose bond type in Mergent FISD is either a government bond (USBD, USNT, USSI, USTC, and USSP), agency bond (FGOV, ADEB and ASPZ, AMTN), mortgage-backed security (MBS), municipal bond (TXMU) or preferred security (PS).

For defaulted bonds, we record and use the return in the month of default (where we exclude accrued interest after default in the return calculation) and exclude the bond thereafter, i.e., we do not include the bond in the portfolio construction if it implies acquiring the bond after the default date but record the return if the bond was acquired before default. If we cannot calculate a return in the default month because the bond does not trade, we substitute in an average default return. The average default return is calculated as the average over non-missing return observations in the default month and the preceding month. The average default return is -9.9%. The rationale for including both the return of the default month and the preceding month is uncertainty about the timing of the default. We use the bankruptcy date as listed in Mergent FISD. If this date is missing then we use the earliest date of a default rating date from Standard&Poors and Moody's.

Additionally, we remove bonds that are not rated, have less than one year to maturity, or has a market value of zero. Finally, we remove transaction prices with a value of 100 in the month after default. The final data set has 1,333,126 bond-month returns.

Considering the fraction of missing returns in TRACE is also of interest as this impacts the ability to calculate returns. Table B.2 shows the average fraction of missing corporate bond returns over the TRACE sample period from July 2002 to December 2021. Panel A

Filter	Deletions	Remaining obs.
Monthly returns from enhanced TRACE		1,947,778
Merge with Mergent FISD	174,940	1,772,838
Exclude variable rate bonds	139,884	1,632,954
Exclude manually detected errors based on 5,000 checks	292	1,632,662
Exclude convertible bonds	68,174	1,564,488
Exclude bonds that do not have a principal value of $1,000	43,846	1,520,642
Exclude trust-preferred securities	5,175	1,515,467
Exclude US corporate PIK bonds	1,232	1,514,235
Exclude US corporate pass thru trust	6,265	1,507,970
Exclude equity-linked bonds	14,999	1,492,971
Exclude non-US non-corporate bonds	2,592	1,490,379
Exclude non-rated bonds	7,199	1,483,180
Exclude bonds with less than one year to maturity	115,255	1,367,925
Exclude market value equal to zero	4,931	1,362,994
Exclude returns beyond the default month	29,868	1,333,126

Table B.1: **Monthly data sample and cleaning.** This table shows how the number of bond-month price observations in our data set is affected by our data-cleaning methods. The initial data is a monthly return series derived from the Enhanced TRACE data set after applying the Dick-Nielsen (2014) error filter, July 2002 to December 2021.

records missing observations at the bond-level (nine digit cusip) and panel B at the firm level (value-weighting bonds with the same permno). A monthly corporate bond return is missing in month t when the bond trades in the previous $t-n$ months but does not trade in month t. In the first column, n is set to 1 while in the second and third column n is fixed at 3 and 6 respectively. Missing percentages are reported both equal- and value-weighted. Equal-weighted percentages are calculated each month by dividing the number of missing returns by the number of bonds trading in the previous $t-n$ months. Value-weighted percentages are calculated by first summing the market values of the bonds with missing returns and then dividing that with the sum of all market values. As seen in the table value-weighting significantly alleviates the impact of missing observations for return calculations.

C Warga and Merrill Lynch/ICE databases

The TRACE database starts in July 2002 and before this date we rely on other data sets based on quotes. In particular, we use so-called Warga data from the Lehman Brothers

	# previous trading months required		
Panel A: Bond-level	1 month	3 months	6 months
Missing (equal-weighted)	20.8%	10.6%	6.1%
Missing (value-weighted)	5.6%	3.2%	2.1%
Avg. bond-month obs.	7474	5535	4561
Panel B: Firm-level	1 month	3 months	6 months
Missing (equal-weighted)	7.5%	4.6%	3.1%
Missing (value-weighted)	0.7%	0.4%	0.3%
Avg. bond-month obs.	871	789	726

Table B.2: Missing corporate bond returns. The table shows the average percentage of missing monthly corporate bond returns and the average number of bond-month observations. The sample period runs from July 2002 to December 2021. In panel A, returns are recorded at the bond-level (nine digit cusip), whilst in panel B, returns recorded at the firm level, value-weighting bonds with the same permno.

Fixed Income database (LBFI) and the Merrill Lynch/ICE (ML) database.

For the period from January 1973 to May 1998, we use monthly data from LBFI. The LBFI database contains monthly prices and returns based on actual quotes and matrix prices. We only use returns based on actual quotes. To have data on bond characteristics, we merge the LBFI database with the Mergent FISD database on bond CUSIP and exclude the same bond types as we do for the TRACE database. Although the LBFI database starts in 1973, the first years consists primarily of matrix prices and as a result our factor returns does not start before February 1985 when data coverage has improved. For each rating category-by-signal group, we require at least 10 observations to compute a factor return. For example, if we don't have at least 10 bonds in the group of SG bonds with a high signal (e.g., long-term reversal), then we compute the SG return as missing (NA). In this case, the factor return is the average of the factor returns with non-missing data (IG+ and IG- in this example). We note that most factors have no rating groups with missing returns, but a few have some rating groups with missing returns before 1991.

We also use ML data for the period from January 1997 to June 2002. This data contains all corporate bonds included in the ML investment-grade and high-yield indices, and we apply the same filtering as for the LBFI database. In the time period in which LBFI and

ML overlap (1/1997–5/1998), we merge these data and use ML when both are available.

The combined data has 851,371 bond-month prices, where 563,363 come from LBFI and 288,008 from ML.

D Additional Empirical Results

Figure D.1 shows the cumulative value-weighted market return using different data filters. The figure shows that there are substantial differences in the market return depending on how the TRACE data is filtered. For example, the annualized excess return (in excess of the one-month Treasury rate) in the period 2002-2021 is 4.86% for our cleaned data and only 4.13% using the JNPS filter. For comparison, the figure also shows the return on the ICE Bank of America US Corporate Total Return Index,[10] which is substantially higher than those using any filtering of TRACE also because the index constituents differ from the filtered TRACE samples.

Figure D.2 shows the firm-level results broken down by rating groups. As was the case with the bond-level analysis, the factors are most significant among IG-, perhaps because this rating group contains a large number of bonds that are more comparable to each other, especially relative to the group of SG bonds, and still have meaningful return differences relative to IG+. Among IG- bonds, the significant factors are those based on amount outstanding, bond age, market value, value, equity momentum, and book-to-price.

As a further robustness analysis, Figure D.3 plots the firm-level factor performance in the early sample 1985-2021 and merged sample 1985-2021. In the early sample, there are only three significant positive factors (market value, equity momentum, credit spread) and a marginally significant negative factor (6m momentum). In the merged sample, there are four significant factors, amount outstanding, bond age, market value, and equity momentum.

Figure D.4 reports the alphas of the 153 corporate bond factors based on equity signals in the bottom panel over the time period 2002-2021. Further, the top panel of the figure shows the performance of average factor returns within each of 13 clusters of factors.

Finally, Figure D.5 reports the posterior mean and confidence intervals for the alphas of

[10] The data is available at https://fred.stlouisfed.org/series/BAMLCC0A0CMTRIV.

the 153 corporate bond factors using the empirical Bayes method of Jensen et al. (2022). The factors are sorted by their posterior mean of the alpha.

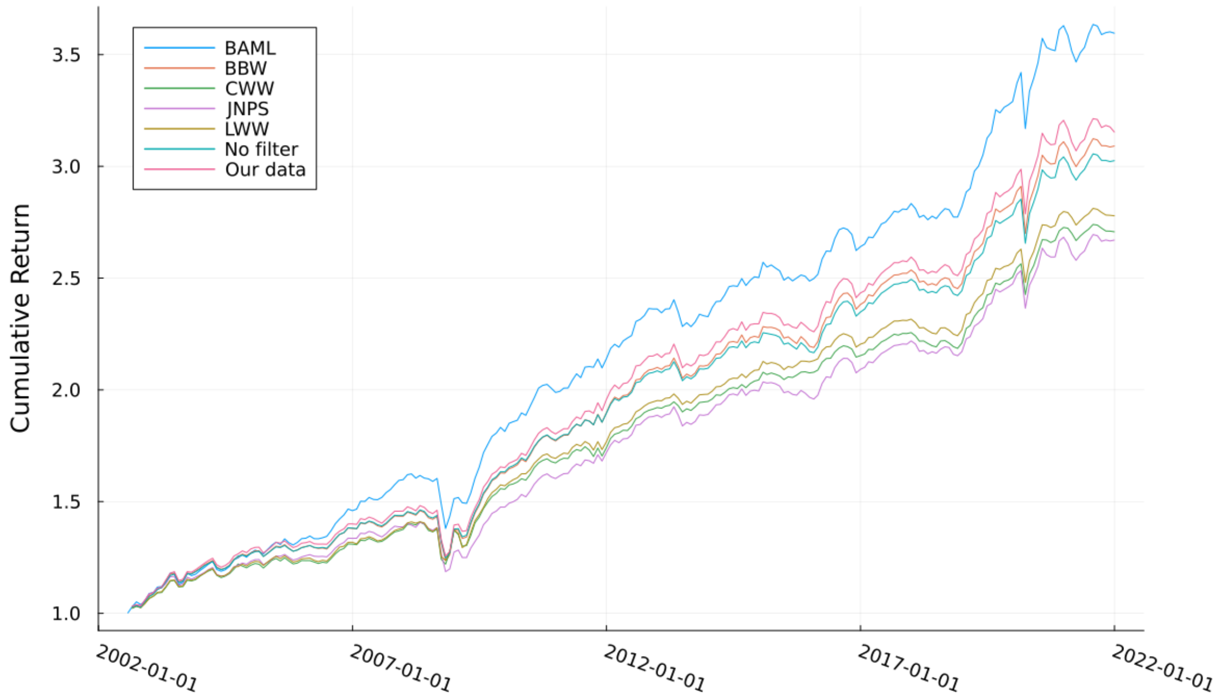

Figure D.1: **Corporate bond market portfolio by data filter.** The figure shows the cumulative market returns with the following data filters: 'BAML' is the Bank of America US Corporate index; 'no filter' is the raw TRACE database after applying the Dick-Nielsen (2014) filter; 'our data' refers to cleaned TRACE data described in Section 1.1; 'BBW' is TRACE data filtered according to Bai et al. (2019); 'JNPS' according to Jostova et al. (2013); 'LWW' according to Lin et al. (2011); and 'CWW' according to Chung et al. (2019). The corresponding average annual returns in excess of the one-month Treasury yield are 5.76% for BAML, 4.94% for BBW, 4.19% for CWW, 4.13% for JNPS, 4.33% for LWW, 4.81% for no filter, and 4.86% for our data. The data period is from August 2002 to December 2021.

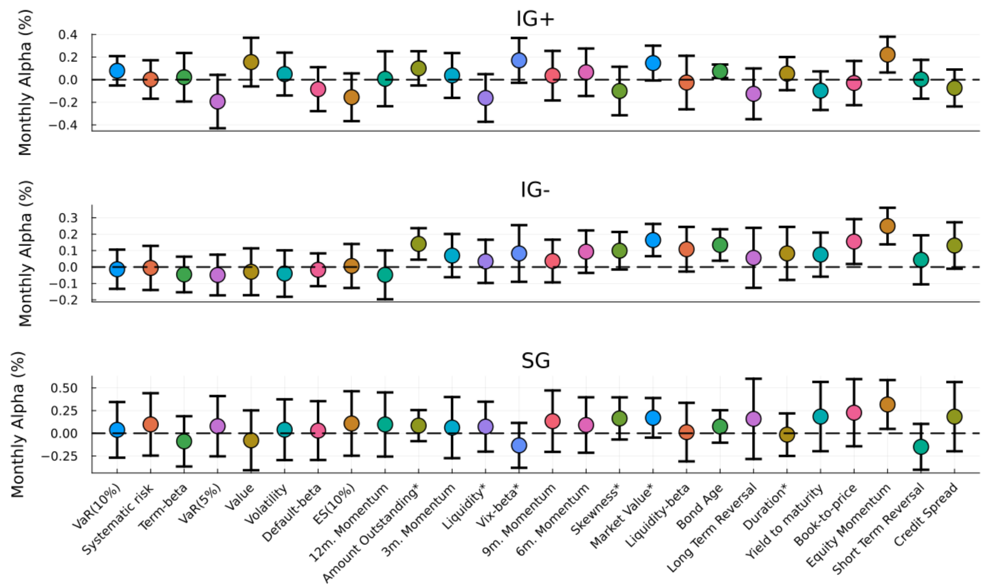

Figure D.2: **Factor performance at the firm level: Breakdown on rating category.** Using factors constructed based on overall firm-level corporate bond returns and the corresponding firm-level characteristics, the figure shows monthly alphas and their confidence intervals in three rating categories: IG+ (top panel), IG- (middle), and SG (bottom). Alpha is the intercept in a regression of monthly excess returns on the corporate bond market factor and the Fama and French (1993) TERM factor, and the sample period is July 2002–December 2021. Factor returns are computed each month from value-weighted corporate bond returns for portfolios that are sorted into three characteristic groups. Each factor is long the high characteristic portfolio and short the low characteristic portfolio, except factors indicated by "*" where the factor is long the low characteristic. Standard errors are Newey-West adjusted.

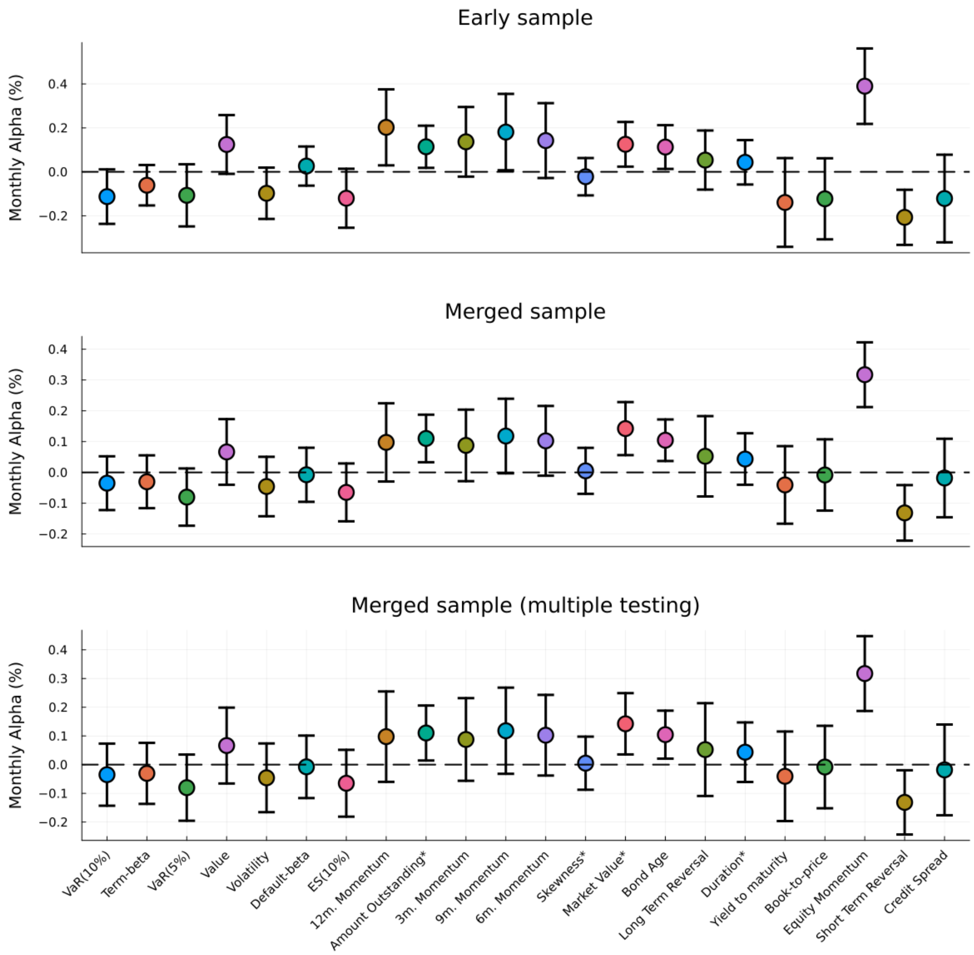

Figure D.3: **Firm-level factors in the early and merged samples.** The figure reports monthly alphas and their confidence intervals for firm-level factors in the early sample 1985-2002 (top panel) and merged sample 1985-2021 (bottom panel). The data from 1985-2021 come from the Warga and LBFI datasets, which are based on quotes, and the data from 2002-2021 comes from TRACE, which is based on trades.

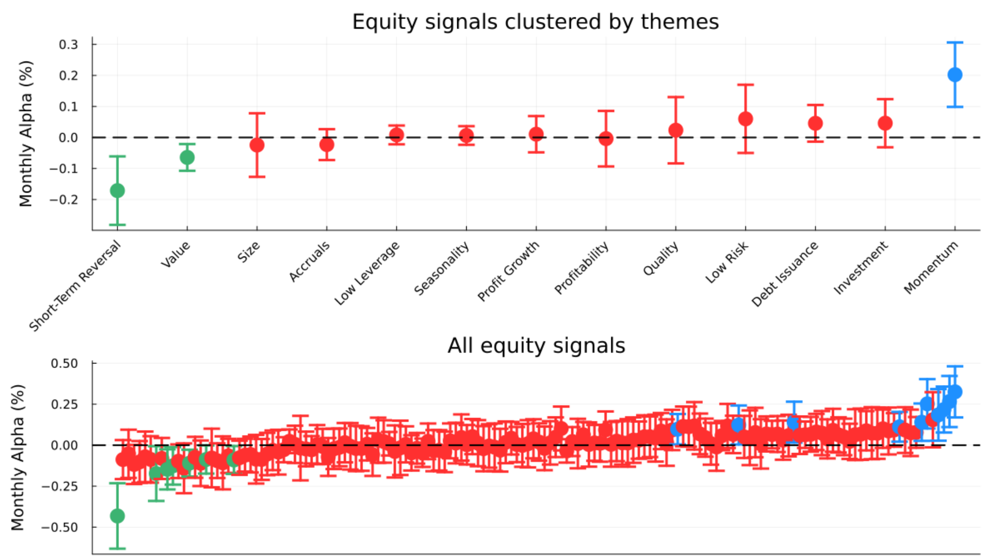

Figure D.4: **Corporate bond factors based on equity signals: TRACE sample 2002-2021 with BH adjustment.** The top panel reports the monthly alphas of the average factor returns within each cluster and their confidence intervals adjusted for multiple-testing effects using Benjamini and Hochberg (1995). Similarly, the bottom panel reports monthly alphas and their MT-adjusted confidence intervals for each of the individual corporate bond factors based on the 153 equity signals from Jensen et al. (2022). As seen in the bottom panel, 10 factors are significantly positive (blue dots), 7 are significantly negative (green dots), and 136 are insignificant (red dots).

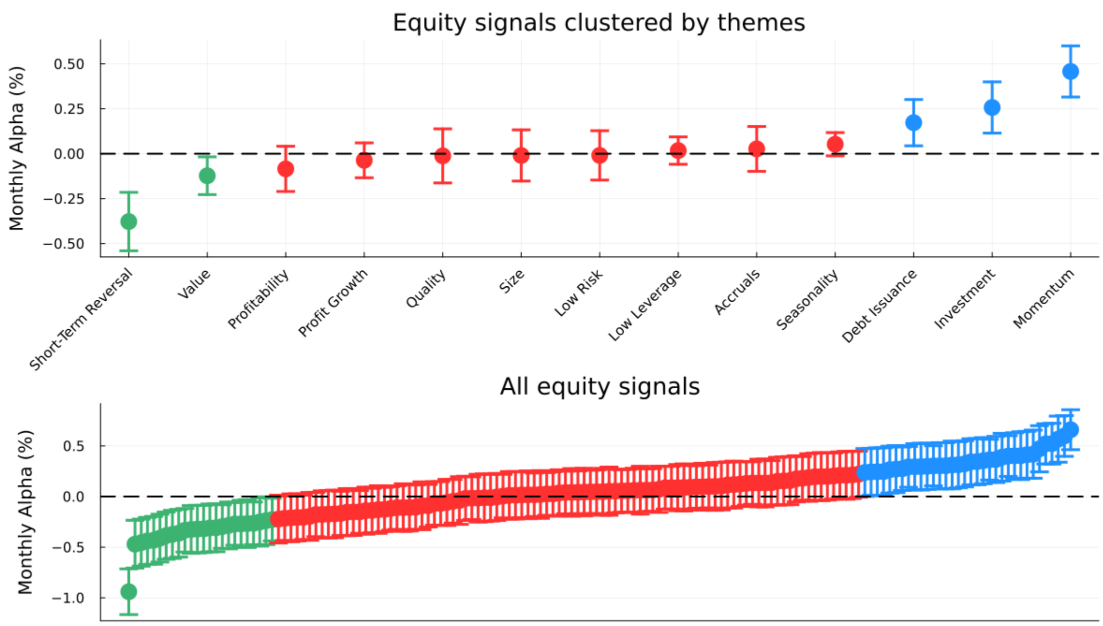

Figure D.5: **Corporate bond factors based on equity signals: Merged sample 1985-2021 with EB-MT adjustment.** The top panel reports the posterior monthly alphas of the average factor returns within each cluster and their confidence intervals adjusted for multiple-testing effects using empirical Bayes estimation. Similarly, the bottom panel reports the posterior monthly alphas and their confidence intervals based on empirical Bayes for the bond factors based on the 153 equity signals from Jensen et al. (2022). As seen in the bottom panel, 34 factors are significantly positive (blue dots), 24 are significantly negative (green dots), and 95 are insignificant (red dots).