Finding value in the U.S. corporate bond market

Liuren Wu*

Zicklin School of Business, Baruch College, The City University of New York

Hashim Zaman†

Laboratory for Innovation Sciences, Harvard University

First draft: January 2, 2023; This version: June 3, 2024

Abstract

This paper identifies value-investing opportunities in the U.S. corporate bond market through the joint construction of a bond valuation model and a return factor model. The valuation model explains the cross-sectional corporate bond yield variation with a flexible functional form in bond risk characteristics including bond duration, credit rating, historical yield change volatility, bond liquidity, and the optionality-induced yield spread adjustment for callable bonds. The return factor model embeds the residual from the valuation model as a mispricing factor while capturing the stronger co-movements between bonds from the same industry, similar rating classes, and similar duration segments, and accounting for differential pricing of bond return risk, liquidity cost, and the optionality exposure. Historical analysis over the past two decades shows that the valuation model can explain the cross-sectional bond yield variation very well, and the value-investing portfolio constructed from the return factor model generates highly positive average excess returns with low risk.

JEL Classification: G11, G12, G13

Keywords: Bond returns; bond yields; credit risk; duration risk; historical volatility; liquidity cost; optionality; mispricing; value investing

---

We thank Linda Allen, Turan Bali, Michael Ning, Amit Sinha, and Georgi Vassilev for helpful discussions. We welcome comments, including references we have inadvertently missed.

*liuren.wu@baruch.cuny.edu.

†hzaman@hbs.edu.

1. Introduction

Systematic value investing in any market relies on two crucial elements: (i) a valuation framework that effectively identifies misvaluation opportunities and (ii) a return factor structure that accurately accounts for the return covariance structure and allows investors to construct well-diversified portfolios targeting the mispricing opportunities without taking on systematic risk exposures. This paper examines the value investing opportunities and the return factor structure in the U.S. corporate bond market.

Based on U.S. corporate bond transaction data over the past two decades, we begin by analyzing the bond yield and return behavior along different risk dimensions. It has been well established since Fama and French (1993) that term risk and credit risk represent two major risk sources of corporate bonds. We show that the average bond yield increases strongly with increasing bond duration and declining credit rating. The higher yield at longer duration reflects the compensation for bearing greater price sensitivity to the underlying yield variation, whereas the yield increase with declining credit rating reflects the risk premium for bearing a greater likelihood of default. The volatility estimates on the ex post bond excess returns over the next month also increase strongly with increasing bond duration and declining credit rating, and the mean bond excess returns increase largely in proportion to the volatility estimates.

Bond duration and credit rating also capture co-movement differences between different bond issuances. We show that the return correlation estimates between bond pairs decline with increasing distance in bond duration and credit rating between the pairs. The declining correlation pattern reflects a multi-dimensional return factor structure along the two risk sources.

In addition to the two structural risk measures, we also construct a historical yield change volatility estimator to capture the interest-rate risk variation not fully reflected in bond duration and credit rating. Even at the same duration and credit rating, the average bond yield level and the ex post return volatility estimates both increase strongly with the historical volatility estimator. The average excess returns also increase in proportion to the return volatility estimates.

As trading volume and volatility vary strongly across different bond issuances, so are the liquidity costs for trading these bonds. We construct a liquidity cost measure for each bond based on its historical yield change volatility and its average trading volume. Bonds with higher liquidity cost have higher average yields, reflecting the well-documented liquidity discount on bond valuation.[1] Bonds with higher liquidity cost also earn higher average excess returns, but the higher average excess return mainly reflects compensation for higher return volatility.

A large proportion of corporate bonds have call provisions (Becker, Campello, Thell, and Yan (2024)). We propose a binomial tree model on the yield of the non-callable component of a callable bond to quantify the value of its call provision. We use the yield spread between the callable bond and its non-callable component to measure the value contribution of the call provision. We also construct an optionality exposure to capture the sensitivity of the embedded call option to the underlying yield volatility variation. We find that the average bond excess return per unit risk increases strongly with the optionality exposure, suggesting that the optionality exposure represents a distinct risk source that demands a higher risk premium per unit risk.

Based on the documented evidence, we propose a bond valuation model to explain the cross-sectional bond yield variation as a function of the bond's major risk characteristics, including bond duration, credit rating, yield-change volatility, liquidity cost, and the option-induced yield spread for callable bonds. We adopt a flexible local linear functional form to allow the dependence structure to vary conditionally with the risk levels. Model estimation shows that the valuation model

[1] See, for example, Houweling, Mentink, and Vorst (2005), Nashikkar and Subrahmanyam (2007), Covitz and Downing (2007), Chen, Lesmond, and Wei (2007), Longstaff, Mithal, and Neis (2005), Bao, Pan, and Wang (2011).

explains the cross-sectional yield variation very well. The cross-sectional R^2 estimates average at 95.65%, and vary within a narrow range between 91.77% and 97.86%. The average explanatory power remains stable across different industry groups, rating classes, and bond durations.

Inspecting the conditional loading coefficient estimates shows that on average, corporate bond yield increases with increasing bond duration, deteriorating credit rating, increasing historical yield change volatility, increasing trading cost, and increasing optionality exposure for bonds with call provisions. However, the relations are not linear, but varying substantially across different regions of the risk characteristics. The dependence on credit rating becomes increasingly stronger as rating deteriorates, but the dependence on duration declines at longer durations, leading to a convex relation with credit rating but a concave relation with duration. The yield dependence on the historical volatility estimator is also very strong and increasingly so at higher volatility levels. By comparison, once controlled for the other risk sources, the liquidity contribution to the yield variation becomes much smaller, similar to literature findings in Longstaff, Mithal, and Neis (2005) and Covitz and Downing (2007).

The valuation model decomposes the yield of each bond into two components: a fair value component reflecting the contribution from its major risk exposures and an idiosyncratic residual component representing the potential mispricing opportunity. We form quintile portfolios based on the residual yield, and show that the different quintile portfolios have similar average risk characteristics, but the average excess return per unit risk increases strongly with the residual yield quintile. The quintile spread portfolio that is long the underpriced fifth quintile and short the overpriced first quintile generates high average excess return with low risk, highlighting the strong predictability of the residual yield on future bond returns.

With the valuation model identifying the mispricing opportunities, we propose a return factor model to capture the bond return covariance structure. The return factor model captures the predictive contribution of the residual yield as a bond mispricing factor while also accounting for the differential pricing of bond return risk, the optionality exposure, and the liquidity cost. In addition, the model includes a multi-dimensional loading matrix to capture the stronger co-movements between bonds from the same industry and similar duration buckets and rating classes.

The return factor model can explain an average of over 40% of the bond excess return variation over the next month. Within the factor structure, the multi-dimensional loading matrix explains an average of 26% of the return variation. It represents a major source of the model's explanatory power, especially during calm market conditions. Meanwhile, the historical return risk factor contributes to an average of 12% of the return variation, and becomes particularly important during market turmoils. The variance contributions from the optionality exposure and the liquidity cost are small and vary strongly over time. By comparison, the mispricing factor explains a small but stable percentage of the return variation across all sample periods.

Under such a return factor structure, the slope coefficient on the mispricing factor represents the future excess return on a statistical arbitrage value-investing portfolio that targets a unit exposure to the bond misvaluation opportunity while being neutral to all systematic risk exposures. The value-investing portfolio generates high average excess returns with low return volatility, leading to extremely high risk-adjusted investment performance. The portfolio generates an information ratio estimate of 4.03 with equal dollar weighting. When we perform weighted estimation on the return factor model accounting for the historical return risk difference and the stronger correlation between bonds from the same issuers, the value-investing portfolio becomes even better diversified, with the information ratio increasing to 7.23.

This paper contributes to the literature in three major areas. First, we propose a decentralized binomial tree model on the yield of the non-callable component of each callable bond to quantify the contribution of the bond's embedded optionality. Unlike instantaneous interest rate tree models (e.g., Hull and White (1996)) that are calibrated to the volatility of a reference interest rate series regardless of which bond the tree prices, our approach fully accounts for the yield volatility difference for different corporate bonds, recognizing that the option value on bonds with higher yield volatilities can be much higher than that computed from a reference interest rate tree.

Second, we propose a statistical valuation model on corporate bonds that can effectively identify bond mispricing opportunities based on the bonds' structural risk characteristics and historical risk behaviors, but without delving into the fundamentals of the issuing company. For public companies, fundamental information from financial statements have been shown to be informative on a company's credit risk and its bond behaviors.[2] Our valuation model takes the average credit rating from rating agencies as inputs, but without analyzing fundamental data directly ourselves, allowing us to analyze the universe of corporate bonds issued by both public and private companies. Meanwhile, compared to simple regression analysis (e.g., Houweling and Van Zundert (2017) and Israel, Palhares, and Richardson (2018)), our valuation model adopts a flexible functional form to capture conditionally varying dependence structures on the risk characteristics and generates very high explanatory powers on the cross-sectional bond yield variation.

Third, we propose a return factor structure that captures not only the return risk, optionality exposure, and liquidity differences across different bond issuances, but also the stronger co-movements between bonds from the same industry, similar rating classes, and similar duration segments. It is well-known that corporate bonds are exposed to credit risk and term risk, but the

[2]See classic statistical and structural models that link a company's default probability to its fundamental information, e.g., Altman (1968), Merton (1974), Black and Cox (1976), Geske (1977), and Leland (1994). Several recent empirical studies explain the bond behavior on public firms with firm fundamental data, e.g., Collin-Dufresne, Goldstein, and Martin (2001), Eom, Helwege, and Huang (2004), Correia, Richardson, and Tuna (2012), Bai and Wu (2016), Kelly, Palhares, and Pruitt (2023), and Wu and Xu (2023).

dimensionalities of the underlying risk sources are less discussed. The literature often uses statistical beta estimates on a term spread and a default spread to capture the risk exposures, or directly uses credit rating and bond duration as risk characteristics, largely reducing each risk source to a single dimension. We show that the cross-correlation estimates between bond returns decline with increasing distances in both bond duration and credit rating, manifesting a high-dimensional risk structure along both the term structure and the credit risk. To capture this behavior, our return factor model embeds a high-dimensional loading matrix to capture the exposures to the different industry groups and to different points on the duration and credit rating spectrum.

Finally, our analysis shows that several bond risk characteristics including duration, credit rating, historical volatility, and liquidity cost can all differentiate bond return risk, but they are not as effective in identifying superior risk-adjusted investment opportunities. The average excess returns on bond portfolios targeting different levels of these risk measures tend to vary largely in proportion to the portfolio's return risk level, leading in small and non-monotonic variations in the information ratio estimates. Through the bond valuation model, we identify highly profitable bond misvaluation opportunities. Embedding such opportunities into the return factor structure produces well-diversified and extremely well-performing value-investing portfolios.

A long list of studies strive to identify "priced" risk factors that can directly predict bond risk premium variations.[3] Among them, Dickerson, Mueller, and Robotti (2023) examine the pricing of a number of commonly proposed corporate bond factors and find them to be statistically and economically insignificant, except for the value-weighted bond market factor. van Binsbergn, Nozawa, and Schwert (2023) highlight the risk premium behavior differences of corporate bonds when the excess returns are constructed over duration-matched Treasury returns instead of short-term Trea-

[3]Prominent examples include Gebhardt, Hvidkjaer, and Swaminathan (2005), Jostova, Nikolova, Philipov, and Stahel (2013), Cao, Goyal, Xiao, and Zhan (2023), van Binsbergn, Nozawa, and Schwert (2023), Dickerson, Mueller, and Robotti (2023), Dang, Hollstein, and Prokopczuk (2023), Dick-Nielsen, Feldhütter, Pedersen, and Stolborg (2023), Kelly, Palhares, and Pruitt (2023), and Dickerson, Robotti, and Rossetti (2024).

sury bill rates. Kelly, Palhares, and Pruitt (2023) employ an instrumented principal components analysis (IPCA) to combine a long list of company and bond characteristics to explain the corporate bond return variation. They show that five principal factors can explain a large proportion of the bond return variation, and that a systematic long-short strategy based on the identified expected returns from the risk factors can generate high risk-adjusted investment returns. Compared to the literature focus on priced risk factors, this paper focuses more on identifying misvaluation opportunities through a bond valuation model construction. Our proposed return factor structure is less for identifying priced risk factors but more for accurately capturing the bond covariance structure so that we can construct well-diversified value investing portfolios without biases in any industry, duration, or credit rating segments, or systematic exposures to any major risk dimensions.

The rest of the paper is organized as follows. Section 2 describes the sample construction procedure and documents the bond yield and return variation along the major risk dimensions. Section 3 proposes a bond valuation model to explain the cross-sectional yield variation as a flexible function of the bond's key risk characteristics. Section 4 proposes a bond return factor model to capture the bond return covariance structure while embedding the yield residual from the valuation model as a mispricing factor. Section 5 concludes.

2. Summary corporate bond behaviors

We obtain data from the Trade Reporting and Compliance Engine (TRACE), a comprehensive source of transaction-level data for U.S. corporate bonds. In January 2001, the Securities and Exchange Commission (SEC) took the initiative to enhance post-trade transparency in the U.S. corporate bond market by mandating the National Association of Security Dealers (NASD) to obtain transaction-level data for publicly issued corporate bonds. This led to the advent of TRACE in July 2002. Since then, bond dealers have reported all trades associated with publicly issued corporate bonds to the NASD, which then makes the trade information public.

2.1. Sample construction

Our sample covers the U.S. corporate bond transactions from July 2002 through March 2022. We obtain bond characteristics such as bond type, maturity date, coupon rate, credit rating, and any embedded optionality information from the Mergent fixed-income securities database (FISD). We merge TRACE data with FISD to construct a comprehensive bond dataset that comprises corporate bond transactions along with bond features.

We apply the following filters to the TRACE data: (1) Restrict the data to bonds of issuers whose country of domicile is the U.S.A. (2) Remove bonds issued through private placement, under the 144A rule, or with non-U.S. dollar denomination. (3) Remove bonds labeled as mortgage-backed, asset-backed, puttable, exchangeable, convertible, or equity-linked. (4) Retain only fixed-rate coupon or zero-coupon bonds with the principal amount of $1,000. (5) Remove transactions that are canceled, corrected, or reversed. (6) Remove transactions identified as when-issued, locked-in, or with a special price indicator. (7) Remove transactions with par value volume less than $1,000. (8) Remove zero-coupon bonds trading above par. (9) Remove transactions that have more than a three-day settlement before September 5, 2017 and more than a two-day settlement after.[4]

Some researchers also filter the trades based on the transaction price or trade size. We only exclude odd lot trades below $1,000 that seem to be outliers in pricing. Historically, there has

[4]On September 5, 2017, the SEC reduced the settlement cycle for U.S. corporate bonds from T + 3 to T+2. Prior to that date, approximately 98% of the trades settled in 3 business days. Once the settlement cycle was reduced, about 95% of the trades in our sample showed settling in 2 business days. On May 28, 2024, the SEC further reduced the settlement cycle to from T+2 to T+1.

been a substantial difference in the pricing available to institutional and retail investors (Edwards, Harris, and Piwowar (2007)). Over the past decade, with the advent of algorithmic trading and the introduction of fixed-income exchange-traded funds (ETFs), there has been an increase in sources of liquidity provision in the U.S. corporate bond market. According to the Securities Industry and Financial Markets Association (SIFMA), as of 2022, about 40% of investment-grade and 30% of high-yield corporate bonds volume are traded electronically. With algorithmic trading, trade size has become much less differentiating of the pricing behaviors or investor types.

TRACE reports the clean price for each transaction, which we use to estimate the daily volume-weighted average price (VWAP) for each bond. The VWAP, in turn, is used to estimate the yield to maturity and monthly returns. We perform valuation analysis on the bond yield once a month, and examine the bond return behaviors over the next month. If a trade exists for a bond in the last 5 business days of a month, we take the VWAP on the last of those 5 days as the month-end price. These month-end prices are used to estimate monthly returns after accounting for accrued interest and coupon payments. Specifically, at each month end t for each bond i, its annualized excess return over the next month is constructed as,

$$r_{i,t+1} = \frac{365}{n_{i,t+1}} \left(\frac{P_{i,t+1} + AI_{i,t+1} + C_{i,t+1}}{P_{i,t} + AI_{i,t}} - 1 \right) - r_{f,t}, \tag{1}$$

where $P_{i,t}$ denotes the clean VWAP price of the bond i at the end of month t, $AI_{i,t}$ denotes the accrued interest, $C_{i,t+1}$ denotes the coupon payment, if any, during the following month $t+1$, and $n_{i,t+1}$ denotes the number of days in between the two month ends $(t, t+1)$. We annualize the monthly return and deduct the one-month Treasury bill rate $r_{f,t}$ to construct the annualized excess return over the next month. The bill rate is obtained from Ken French's data library.

In addition to the pricing information, we also obtain the industry classification and credit rating information for each issue from Mergent FISD. Most of the bond issuances are classified into three broad industry groups: industrials, financials, and utilities. We exclude bond issuances that are either without the industry information or fall out the three industry classifications.

We make use of credit ratings assigned by both S&P and Moody's, the two largest and most widely followed rating agencies in the U.S. Each letter rating is assigned a numeric value ranging from 1 for AAA, 2 for AA+, to 21 for C. For any issue at a point in time, if ratings from both agencies are available, we assign the mean of the two ratings to the issue. If only one of the two ratings is available, that credit rating is assigned to the issue. We filter out bond issuances without credit rating information.

Finally, we exclude bonds from our analysis that do not have enough trading information for the construction of the next month return, bonds with less than one year to maturity, and bonds with negative yields or yields greater than 30%. Bonds with extremely high yields are often in the process of going through financial distress, adding great uncertainty to trading execution and investment. Even with this filtering, our selected sample still includes 82 observations where the bonds are defaulted or defeased during the following month. For these observations, we approximate the next month return by assuming a 40% recovery.

The filtering process generates a total of 1,224,423 bond-month observations, including 36,102 bonds issued by 3,814 unique issuers from July 2002 to March 2022. The total number of selected bonds at each month ranges from a low of 2,936 to a high of 6,666 with an average of 5,188 bonds.

Out of the pooled sample, 51% are from industrials, 38% from financials, and 11% from utilities. Figure 1 plots the pooled distribution of the bond rating in panel A, estimated on the numeric rating with a Gaussian kernel. The distribution peaks around 8, corresponding to BBB+ rating. Panel B plots the pooled distribution of the bond duration, which represents the value-weighted average maturity of the bond's cash flows. The duration distribution is heavily skewed with a long right tail. The distribution peaks around 4 years, with a secondary peak around 12 years. The maximum duration reaches 32 years.

[Figure 1 about here.]

With the constructed sample, we document the summary behaviors of corporate bond yields and bond excess returns along their major risk dimensions. The documented evidence serves as the starting point for developing the corporate bond valuation model and the return factor structure in the latter sections. We consider both structural risk characteristics such as bond duration and credit rating, as well as statistical volatility and liquidity cost estimates. In addition, as a large proportion of the corporate bonds include call provisions, we develop an algorithm to quantify the value contribution and the optionality risk exposure of the embedded call provision and examine their impacts on bond yield and bond excess returns.

2.2. Bond behavior variation across duration and credit rating

Term risk and credit risk represent two major structural risk sources for corporate bonds. To examine how the corporate bond yield and return behavior vary across different maturity buckets and rating classes, at each month, we construct bond portfolios targeting different maturity levels and credit rating classes. We use the bond duration (m) as a measure of the value weighted maturity of each bond's cash flows and we use the numeric values from 1 to 21 to represent the credit rating (c) variation from AAA to C. For each portfolio k that targets a certain duration and credit level

(m_k, c_k), we construct the weighting on each bond i in the universe with a bivariate Gaussian kernel,

$$w_{ik} = \exp\left(-\frac{(\ln m_i - \ln m_k)^2}{2h_m^2} - \frac{(c_i - c_k)^2}{2h_c^2}\right), \tag{2}$$

where (h_m, h_c) denote the bandwidths that control the speed of weighting decay with the bond's distances to the target duration and numeric rating, respectively. We set the bandwidth on the numeric rating to $h_c = 3$ to cover a full letter grade. As the duration distribution is heavily skewed to the right and the information along the term structure tends to dissipate exponentially (Calvet, Fisher, and Wu (2018)), we define the duration distance in the log space and set the bandwidth to $h_m = \ln 1.5$, corresponding to a duration distance ratio of 1.5. We normalize the weighting on each portfolio so that the portfolio weights sum to one dollar. At each month, for each bond portfolio, we compute its average bond yield and its annualized excess return over the next month.

Table 1 reports the portfolio's average yield and the summary statistics of the portfolio's excess returns over the next month, including the annualized mean excess return (Mean), the annualized return volatility (Volatility), and the annualized information ratio (IR) constructed as the ratio of the mean excess return to the return volatility.

[Table 1 about here.]

Panel A reports the bond portfolio behaviors that target different durations while holding the target credit rating fixed at BBB+, the mode of the pooled credit rating distribution. Duration measures not only the average maturity of the bond's cash flows, but also the magnitude of bond return variation per unit variation in the bond's yield, i.e., the bond's interest-rate sensitivity. The average bond yield increases strongly with duration. This positive average slope against bond duration is often referred to as the term risk premium, reflecting the compensation to longer-term bondholders for bearing greater price sensitivity to interest-rate changes.

Because of the increased interest-rate sensitivity, the ex post bond return volatility estimates increase strongly with the target duration. The mean excess return of the portfolio also increases strongly with duration, largely in proportion to the increase in the return volatility estimates. The information ratio estimates do not vary monotonically with the target duration. They are moderately higher at intermediate maturities, and lower at the two ends of the duration spectrum.

Panel B reports the bond portfolio behaviors that target different credit rating while holding the target duration fixed at four years at the mode of the pooled duration distribution. The average bond yield increases strongly with deteriorating credit rating. The positive yield spread between low and high credit rating classes is often referred to as the credit risk or default risk premium, reflecting compensation for bearing a greater likelihood of default.

The bond return volatility estimates increase strongly with declining credit rating, suggesting that the bond portfolios with higher credit risk also experience larger return volatility. The mean excess returns of the portfolios again increase largely in proportion to the return volatility estimates. Historically, researchers have noticed that bonds who had recently fallen from investment grade to high yield rating, i.e., "fallen angles," tend to generate higher risk-adjusted returns on average due to forced selling activities (Ambrose, Cai, and Helwege (2008)). Potentially related to this finding, the information ratio estimates in panel B are the highest for portfolios targeting ratings BBB to BB, but become somewhat lower at the two ends of the rating spectrum.

In addition to differentiating bond yield and return risk, duration and credit rating can also differentiate co-movements between different bond issuances. Carr and Wu (2023) show that the correlations between changes in swap rates decline with the absolute distance in swap maturities. We anticipate similar correlation dependences of corporate bond returns on the duration distances between the bond issuances. To examine this hypothesis, we measure the cross-correlations of the excess return series between the bond portfolios targeting different duration levels at the fixed rating class of BBB+. Figure 2 plots in panel A the pairwise correlation estimates as a function of the absolute distance in the natural logarithm of the target durations of the portfolios. At zero duration distance, the correlation of a return series is 100% with itself. As the duration distance increases, the correlation estimates decline. In the plot, we have created bond portfolios targeting durations from one to 32 years, covering the full duration range of the data sample. The maximum distance is between the 32-year duration portfolio and the one-year duration portfolio, with a log duration distance of 3.5 times. The two return series have the lowest correlation estimate at 55.5%.

[Figure 2 about here.]

We hypothesize that credit rating can also differentiate co-movements between bond issuances. To examine this hypothesis, we measure the return correlation between bond portfolios at different rating classes targeting a fixed four-year duration. Panel B of Figure 2 plots the correlation estimates against the absolute numeric rating distance. The correlation estimates decline sharply as the numeric rating distance increases. At the largest distance of 20 between the bond portfolio targeting a numeric rating of 21 and the portfolio targeting a numeric rating of 1, the correlation estimate between the two return series is merely 37.8%.

It is well-known that corporate bonds are exposed to credit risk and term risk; however, the dimensionalities of the underlying risk sources are less discussed. Researchers often use beta estimates to a term premium and a default premium to capture the risk exposures, or directly use the numeric credit rating to proxy the credit risk exposure and duration to proxy the term risk exposure. In either case, the practice largely reduces each risk source to a single dimension. Yet, the term structure literature has long recognized the multi-dimensional nature of the interest rate term structure variation (e.g., Litterman and Scheinkman (1991) and Dai and Singleton (2000)). The fact that the correlation estimates decline with increasing duration distance is a manifestation of such a multi-factor structure. Similarly, the fact that the correlation estimates decline with increasing rating distance also implies a multi-factor structure along the credit rating dimension. In building a corporate bond return factor model, one must fully account for the high-dimensionalities along the term and credit risk dimension to accurately capture the return covariance structure.

2.3. Historical interest-rate volatility

Duration and credit rating represent structural measures of bond risk. We can also directly measure the risk behavior of a bond based on its historical behaviors. At each month t and for each bond issuance i, we construct a historical volatility estimator ($v_{t,i}$) on its monthly yield changes with a 12-month historical rolling window. For issuances without a full 12-month history, we pool the historical yield changes over the past 12 months over the whole sample and construct a weighted variance estimator targeting the duration, rating, and yield level of the bond issuance in question. The pooled estimator puts more weight to observations (j) closer to the target bond issuance (i) in terms of duration, rating, and yield level via a trivariate Gaussian kernel,

$$w_{ji} = \exp\left(-\frac{(\ln m_j - \ln m_i)^2}{2h_m^2} - \frac{(c_j - c_i)^2}{2h_c^2} - \frac{(y_j - y_i)^2}{2h_y^2}\right), \tag{3}$$

where the bandwidths on duration and rating distances are set the same as before and the bandwidth on the yield distance is set to default choice (Simonoff 1996). We take a weighted average of the bond's own estimator and the pooled estimator, where the weight on its own estimator increases linearly with the number of historical yield change observations on the bond over the past 12 months. The weight reaches 100% on its own estimator when the bond has a full 12-month history.

The pooling estimation resolves the lack of history issue for a large proportion of the bonds.

To examine whether the volatility estimator can capture different bond behaviors beyond the structural risk characterization with duration and credit rating, we form bond portfolios targeting different percentile values of the volatility estimator at the 10, 25, 50, 75, and 90th percentile while holding the target duration fixed at four year ($\overline{m} = 4$) and the target rating fixed at BBB+ ($\overline{c} = 8$). For each target volatility percentile k, we construct the portfolio weight on each bond i according to the following trivariate Gaussian kernel,

$$w_{ik} = \exp\left(-\frac{(\ln v_i - \ln v_k)^2}{2h_v^2} - \frac{(\ln m_i - \ln \overline{m})^2}{2h_m^2} - \frac{(c_i - \overline{c})^2}{2h_c^2}\right), \tag{4}$$

where the bandwidths on duration and rating distances are set the same as before and the bandwidth on the volatility (h_v) distance is set to default choice. The distribution of the volatility estimator has a large right tail. We measure the volatility distance in the log space.

Table 2 summarizes the bond portfolio behavior in panel A. The time-series average of the percentile values of the volatility estimator increases from 0.85% at the 10th percentile to 4.12% at the 90th percentile. The average yield of the portfolio increases strongly from 3.45% at the 10th percentile to 5.8% at the 90th percentile. Therefore, even within the same rating class and the same duration bucket, the historical volatility estimator can further differentiate the bond yield behavior. Bonds with higher historical volatility estimators tend to have higher yield levels.

The remaining columns of Table 2 report the summary statistics on the portfolio's excess returns over the next month. The ex post return volatility estimates of the bond portfolios tend to increase with the ex ante historical yield change volatility estimates. The mean excess returns of the bond portfolios increase largely in proportion to the return volatility estimates. The information ratio estimates do not vary strongly or monotonically with the yield change volatility quintile.

2.4. Liquidity cost

The U.S. corporate bond market has experienced tremendous growth during the past two decades. According to statistics from SIFMA, the size of the market has doubled from \$5.9 trillion in 2008 to \$10.8 trillion in 2023. The average daily trading volume has grown from \$14 billion in 2008 to over \$50 billion in 2024. Despite the rapid growth, the corporate bond market is still considered illiquid relative to its equity and treasury counterparts. Not only do the trading activities vary greatly across different bond issuances, but the market as whole has shown vulnerabilities during time of crises when the market volatility spikes up (Lester (2021), Kargar, Lester, Lindsay, Liu, Weill, and Zúñiga (2021)). This section examines how the liquidity of a bond issuance affects its yield and return behavior.

The literature has proposed many different liquidity measures for a financial security. Fundamentally, the illiquidity of a security represents the cost of trading a certain amount of the security, or equivalently the risk the liquidity provider faces during the time it takes to unload the position. The risk increases with the volatility of the security, but declines with the trading volume because it takes a shorter period of time to unload a position when the trading volume is higher. If one adopts a volume-weighted trading (VWAP) strategy that takes a fixed proportion h of the market volume, the time it takes to unload a certain size becomes proportional to the trade size as a fraction of the market trading volume. Therefore, for a given trade size, a commonly used trading cost measure is constructed to be proportional to the ratio of the security's volatility forecast ($v_{t,i}$) to the square root of the security's daily trading volume forecast ($u_{t,i}$) (Grinold and Kahn 1999),

$$q_{t,i} = v_{t,i}/\sqrt{u_{t,i}}. \tag{5}$$

If we represent the daily trading volume in million dollars, $1/(hu_{t,i})$ represents the fraction of a day it takes to unload a million dollars of the security if we follow a VWAP trading strategy that takes a fixed fraction h of the market volume. As volatility scales with the square root of time, the ratio in (5) captures the risk the liquidity provider faces during the time it unloads the position. In general, the liquidity provider charges a premium in proportion to the risk she faces, whether in terms of commission, bid-ask spread, or market impact. The measure in (5) is similar to the Amihud (2002) liquidity measure, but with better risk interpretation and better matching of empirical evidence on market impacts (Gatheral and Schied (2013)).

We construct the liquidity cost measure in (5) for each bond issuance at each month based on the historical yield-change volatility estimator constructed in the previous section and the bond's average daily trading volume over the past 12 month. We form bond portfolios targeting different percentile values of the liquidity cost measure at the 10, 25, 50, 75, and 90th percentiles while holding the target duration fixed at four year ($\overline{m} = 4$) and the target credit rating fixed at BBB+ ($\overline{c} = 8$). For each target liquidity cost percentile k, we construct the portfolio weight on each bond i according to the following trivariate Gaussian kernel,

$$w_{ik} = \exp\left(-\frac{(\ln q_i - \ln q_k)^2}{2h_q^2} - \frac{(\ln m_i - \ln \overline{m})^2}{2h_m^2} - \frac{(c_i - \overline{c})^2}{2h_c^2}\right), \tag{6}$$

where the bandwidths on duration and rating distances are set the same as before and the bandwidth on the volatility (h_q) distance is set to default choice. Similar to the weighting on volatility in (14), we measure the liquidity cost distance in the log space.

Panel B of Table 2 summarizes the bond portfolio behavior at different liquidity cost percentiles. The first column reports the time-series averages of the liquidity cost estimate (q). In constructing the measure, we represent the volatility estimator in annualized percentage points and the average daily trading volume in million dollars. The average liquidity cost estimates increase from 0.35 at the 10th percentile to 7 at the 90th percentile. The liquidity cost increase can come from an increase in volatility and/or a decline in trading volume.

As the liquidity cost increases, the average bond yield increases from 3.57% at the 10th percentile to 4.83% at the 90th percentile, reflecting the well-documented liquidity discount on bond valuation. Furthermore, since the liquidity cost increase partially reflects an increase in yield change volatility, the ex post return volatility estimators also become higher for portfolios with higher liquidity cost. The average excess returns increase with the increased return volatility.

In the literature, several studies examine whether liquidity shocks are priced in corporate bond returns.[5] The statistics in panel B of Table 2 show that illiquid bonds are riskier and the average excess returns are higher due to the increased risk; nevertheless, the average excess return per unit risk, as measured by the information ratio, does not vary strongly or monotonically with liquidity.

2.5. Call provisions

A large proportion of corporate bond issuances are callable. The call provisions allow firms to take advantage of early refinancing opportunities by calling existing bonds and issuing new debt as yields drop.[6] They can also help firms manage debt maturity (Xu (2018)), mitigate rollover risk (Ma, Streitz, and Tourre (2023)), and prevent debt overhang in corporate mergers (Becker, Campello, Thell, and Yan (2024)). The call provisions affect both the bond valuation and the bond risk behavior in profound ways. First, as a callable bond can be decomposed into a non-callable bond minus a call option, the call option reduces the bond value and increases the bond yield. Second, the convex call payoff structure induces exposures to interest-rate volatility movement. This

[5]See for example, De Jong and Driessen (2012), Acharya, Amihud, and Bharath (2013), Lin, Wang, and Wu (2011), Downing, Underwood, and Xing (2005), Dick-Nielsen, Feldhütter, and Lando (2012).

[6]See, for example, Merton (1974); Brennan and Schwartz (1977); Vu (1986); Mauer (1993); Longstaff and Tuckman (1994); Acharya and Carpenter (2002); Guntay, Prabhala, and Unal (2004); Jarrow, Li, Liu, and Wu (2010).

optionality exposure represents a distinct risk source that can be priced differently from interest-rate and credit risk.

To account for the optionality of a callable bond, the industry often starts with a reference interest-rate curve stripped from either Treasury bonds or US dollar LIBOR/swap rates. An instantaneous interest rate tree, e.g., Hull and White (1996), is constructed based on the volatility estimate on the reference interest rate, and an option-adjusted spread is added as a parallel shift of the reference interest rate curve to price a callable bond. This approach can work well with agency bonds with small default probabilities, but it fails to account for the much higher yield volatility for corporate bonds, especially for bonds at low rating classes.

In this section, we propose a binomial tree directly calibrated to the yield of the non-callable component of a callable corporate bond, fully accounting for the yield volatility differences for different bonds. Formally, to price a callable bond k at time t, we use $y^n_{t,k}$ to denote the yield on the non-callable component of this bond, and $v_{t,k}$ the annualized volatility on the yield change. Let (i,j) denote the node of the binomial tree on this non-callable bond yield at the ith time step and the jth node from top to bottom. Starting at node $(0,0)$ with the non-callable yield being $y^n_{t,k}$ for this particular bond, the yield at the (i,j) node is constructed as

$$y_{i,j} = y^n_{t,k} + (i-2j)v_{t,k}\sqrt{\Delta t}, \quad j = 0,1,\cdots,i, \tag{7}$$

where Δt denotes the time step of the tree. The recombining tree has a step size of $v_{t,k}\sqrt{\Delta t}$ and generates $(i+1)$ nodes at the ith time step.

Since we only observe the callable bond price and not its non-callable component, we construct the volatility estimator by pooling yield changes of all non-callable bonds during the past 12 months and applying the bivariate Gaussian kernel weighting in (2) to generate a weighted variance estimator targeting the duration and credit rating of the bond in question.

At each node (i, j), we set its branching probability such that the present value of the future cash flows from the non-callable bond at node (i, j) is equal to the probability-weighted present values of the same future cash flows at its two branches discounted back by the yield at (i, j),

$$PV_i(y_{i,j}) = e^{-y_{i,j}\Delta t}\left(PV_{i+1}(y_{i+1,j})p_{i,j} + PV_{i+1}(y_{i+1,j+1})(1-p_{i,j})\right), \tag{8}$$

where $p_{i,j}$ denotes the probability going from node (i, j) to its upper branch $(i+1, j)$ and $PV_i(y)$ denotes the time-i present value of the bond's future cash flows evaluated on the yield y. When the volatility estimate is large, the yield tree can become negative, at which node we value the bond at zero yield and adjust the branching probability accordingly.

Compared to an interest-rate tree built on a reference interest-rate curve, which can be used to price bonds with different maturities, coupons, and even different risk profiles with a parallel shift of an option-adjusted spread, the yield tree in (7) is constructed specifically to price the particular bond k at time t. The step size of the tree matches the estimated volatility level of the historical yield changes of bonds at similar duration and rating classes. The branching probability is constructed to match the pricing of the cash flows on this particular bond. We need to construct a separate yield tree for each bond based on its own future cash flows and its own historical risk behavior. This decentralized approach, similar to the decentralizing bond pricing of Carr and Wu (2023), allows us to match the behavior of each bond without being tied to any particular interest-rate curve.

The call provisions of corporate bonds can take on different forms: make-whole, hybrid, or fixed-price. Make-whole bonds are callable during the entire make-whole period, which typically starts at or just after issuance and ends at or just before maturity (Powers and Sarkar (2013)). The exercise price is determined by discounting the remaining cash flows of the bond at the treasury rate plus a small make-whole spread, which is typically 50 basis points or 15% of the at-issue credit spread (Mann and Powers (2003b)). As the make-whole exercise price tends to be much higher than the prevailing market price of a corresponding non-callable bond, it is rarely optimal for the issuer to exercise the make-whole call provision from a purely pricing perspective. We hence ignore the optionality of the make-whole provision. The hybrid bonds are structured as make-whole bonds in the early part of their life and are then converted to fixed-price callable bonds, typically after 50% of their life. The fixed-price bonds are callable based on a fixed schedule of call dates and prices determined at the time of issuance (Powers (2021)).

The callable bonds can be called at different frequencies. A majority of them (88.6%) can be called continuously within a specified call date range. The call dates can contain multiple entries corresponding to call strike price changes. When the provision provides a single call date and a single call strike, but with the frequency specified as continuous, the bond can be called at any date between the listed single call date and the bond's expiry at the same exercise price. Other call frequency specifications include daily, monthly, quarterly, semi-annual, annual, every 2, 3, 4, or 5 years, and just once. We construct the tree with a time interval of half a month ($\Delta t = 1/24$). Both daily and continuous call provisions are approximated with discrete call decision at the half-month frequency.

We assume frictionless market and optimal call exercising, and set the callable bond value equal to the non-callable bond value minus the call option value. The pricing involves an iterative procedure: We start by setting the call option value to zero and setting the non-callable bond yield to the callable bond yield. Under this initial guess, we construct the non-callable yield tree and compute the call option. Then, we add the computed option value to the callable bond price to generate a new non-callable bond price and a non-callable yield. We reconstruct the yield tree based on the new yield estimate and iterate until convergence.

Out of the 1,224,423 month-bond observations, 922,385, or 75% of them are flagged as callable. Among them, 503,044 have fixed call strike prices and call dates for us to perform the option valuation, constituting 41% of the sample. The rest of the callable bonds are predominantly of the purely make-whole type, which we assume zero optionality.

For each callable bond, once we compute its option value with the binomial tree, we estimate the value contribution of the call provision with an option-induced yield spread,

$$s_{t,k} = y_{t,k} - y^n_{t,k}, \tag{9}$$

with y being the callable bond yield and y^n the yield of the corresponding non-callable component. As the callable bond is valued as the non-callable bond minus the call option, a positive call option value reduces the value of the callable bond and increases its yield. The yield spread captures the call option's value contribution in the yield space.

The convex payoff structure of the call option also induces exposures to interest-rate volatility variation. We estimate the volatility sensitivity of the call option by shocking the volatility input of the tree while keeping the yield of the non-callable bond component fixed,

$$g_{t,k} = \frac{c(v_{t,k}+\varepsilon) - c(v_{t,k}-\varepsilon)}{2\varepsilon}\ , \tag{10}$$

where $c(v_{t,k})$ denotes the call option value as a function of the volatility input $v_{t,k}$ and ε denotes a small volatility shock. We multiply the volatility sensitivity estimate ($g_{t,k}$) by the yield-change volatility estimator ($v_{t,k}$) to approximate the different magnitude of the volatility shocks for different bonds. We then scale the exposure by the callable bond price ($P_{t,k}$) to generate an optionality exposure measure ($o_{t,k}$) matching the scale of the bond return construction

$$o_{t,k} = \frac{g_{t,k} v_{t,k}}{P_{t,k}}. \tag{11}$$

To examine how the call provision affects the bond valuation and the ex post bond return behavior, at each month we form quintile portfolios based on (i) the option-induced yield spread $s_{t,k}$ and (ii) the optionality exposure $o_{t,k}$. The yield spread represents the bond value contribution in the yield space, whereas the optionality exposure captures the option's return risk contribution. Table 3 summarizes the quintile portfolio behaviors constructed based on both measures. The quintile portfolios are formed within the subsample of callable bonds with fixed strike prices.

[Table 3 about here.]

The left side of the table under panel (i) reports the quintile portfolio behaviors constructed based on the option-induced yield spread. Panel A reports the average yield spread and optionality exposure of the quintile portfolios. The average yield spread increases from 44 basis points in the first quintile to 2.77% in the fifth quintile. While the literature has repeatedly documented the average yield differences between callable and non-callable bond,[7] our binomial tree estimation shows that depending on the particular structure of the call provision and the volatility of the underlying yield, the option-induced yield spread can vary greatly across the different callable bonds. As the value contribution increases through the quintiles, the average optionality exposure also shows a moderate increase.

Panel B reports the time-series averages of the quintile portfolio's summary characteristics,

[7] See, for example, Kish and Livingston (1993); Mann and Powers (2003a); Jameson, King, and Prevost (2021); Mann and Powers (2003b); Mann and Powers (2003a); Powers and Tsyplakov (2008); Park and Clark (2015); and Jameson, King, and Prevost (2021).

including the average rating, duration, and yield level. The average numeric rating increases with the quintile whereas the average bond duration declines. The average yield level stays roughly the same across the quintiles, even though a larger proportion of the yield in the higher quintiles comes from the option value contribution.

Panel C reports the summary behaviors of the quintile portfolio's excess returns over the next month. The mean excess return declines with the quintile, so does the return volatility estimates. As the yield spread increases across the quintiles, the call value becomes larger and the call option is likely to become closer to be in the money and to be called. As such, the effective duration, and accordingly the exposure to interest-rate and credit risk, becomes smaller, leading to lower return volatility estimates. Nevertheless, although the yield spread captures the value contribution of the call provision, it is not an effective measure of the risk exposure induced by the call provisions. The cross-quintile variation of the information ratio estimates does not show any clear pattern.

The right side of the table under panel (ii) summarizes the behavior of the quintile portfolios constructed on the optionality exposure to the interest-rate volatility variation. The average optionality exposure increases from 0.27% at the first quintile to 1.08% at the fifth quintile. The yield spread does not vary monotonically with the optionality exposure, highlighting the fact that the variation of the value contribution can be quite different from the optionality exposure. In general, the value of a call option is higher when it is more in the money, whereas the volatility sensitivity, or the optionality exposure, is high when the option maturity is long and the option is close to being at the money.

Both the average numeric rating and the average bond duration increase with the optionality exposure. The bond yield also becomes higher at the higher quintile due to the higher credit and term risk. Most striking is the quintile portfolio's excess return behavior. As the optionality exposure increases, the mean excess return increases strongly from -0.26% to 10.15% per annum. The average excess return per unit risk, or the information ratio, also increases strongly with the optionality exposure, from -0.03 at the first quintile to 0.7 at the fifth quintile. Therefore, different from term risk, credit risk, and the statistical volatility and liquidity cost estimates, which can all differentiate bond return risk but cannot differentiate the average excess return per unit risk in a monotonic pattern, the information ratio estimates show a strong monotonically increasing pattern with the optionality exposure.

The fact that the average excess return per unit risk increases monotonically with the optionality exposure suggests that the optionality exposure, or the volatility risk, is indeed priced differently from the interest-rate risk and the credit risk. In particular, the embedded short optionality exposure in the callable bonds generates a higher risk premium per unit risk, in line with the general observation that investors are willing to pay for positive optionality.[8]

3. A yield model of bond valuation

Based on the evidence documented in the previous section, we propose a corporate bond valuation model, which represents the value of a corporate bond in the yield space and models the bond yield as a flexible function of its main risk characteristics,

$$\mathbf{y}_t = f_t(\mathbf{c}_t, \mathbf{m}_t, \mathbf{v}_t, \mathbf{q}_t, \mathbf{s}_t) + \mathbf{e}_t, \tag{12}$$

[8]The most direct evidence is the observed negative variance risk premium in the stock index options market (Carr and Wu (2009)). Through a joint analysis of stocks and corporate bonds within a structural framework, Wu and Xu (2023) identify a strongly negative risk premium in the embedded optionality of stocks and bonds, the absolute information ratio of which is higher than that of the directional firm risk exposure. Several studies also find that stocks with lottery-like payoffs are priced more expensively, e.g., Barberis and Huang (2008), Kumar (2009), Conrad, Kapadia, and Xing (2014), and Eraker and Ready (2015).

where $\mathbf{y}_t$ denotes an $(N_t \times 1)$ vector of bond yields at month t with N_t being the size of the bond universe at time t. The model explains the cross-sectional variation of the corporate bond yields with the corresponding variations in the major risk dimensions: credit rating $\mathbf{c}_t$, duration $\mathbf{m}_t$, historical yield change volatility $\mathbf{v}_t$, liquidity cost $\mathbf{q}_t$, and the option-induced yield spread $\mathbf{s}_t$ for callable bonds. Rating and duration capture the structural risk characteristics of a bond. The historical volatility estimator captures the interest-rate risk of a bond not fully reflected in its duration and credit rating. The liquidity cost measure accounts for the idiosyncratic liquidity difference of the different issuances. For callable bonds, we estimate a yield spread induced by the embedded call provision through a binomial tree model. Adjusting for this option-induced yield spread makes the bond yields more comparable across bonds with different call value contributions.

We apply a flexible function form $f_t(\mathbf{c}_t, \mathbf{m}_t, \mathbf{v}_t, \mathbf{s}_t)$ to accommodate the potentially nonlinear dependence structures on the risk characteristics. Specifically, at each month t, for non-callable and pure make-whole bonds, we estimate a local linear cross-sectional relation between the bond yields and their numeric rating, duration, volatility, and liquidity cost estimates,

$$\mathbf{y}_t = [\mathbf{1}, \mathbf{c}_t, \mathbf{m}_t, \mathbf{v}_t, \mathbf{q}_t]\, \mathbf{b}(\mathbf{c}_t, \mathbf{m}_t, \mathbf{v}_t, \mathbf{q}_t) + \mathbf{e}_t. \tag{13}$$

The relation is estimated repeatedly for each bond k with a weighted regression, where the weighting on each data point i is constructed with a multivariate independent Gaussian kernel,

$$w_{ik}^n = \exp\left(-\frac{(c_i - c_k)^2}{2h_c^2} - \frac{(\ln m_i - \ln m_k)^2}{2h_m^2} - \frac{(\ln v_i - \ln v_k)^2}{2h_v^2} - \frac{(\ln q_i - \ln q_k)^2}{2h_q^2}\right). \tag{14}$$

The distributions of the volatility and liquidity cost estimators have long right tails. To increase the numeric stability for the regression at the sparsely populated right tail region, we winsorize the volatility and liquidity cost estimates at their 99th percentile and build their distances in the log space. We set the bandwidth on rating and duration the same as before, and the bandwidths on the log volatility and liquidity cost distance at three times the default choice. A smaller bandwidth choice allows the model to be more tuned to the local features whereas a larger bandwidth smoothes out more of the data noise. Our relatively conservative bandwidth choice achieves reasonably good pricing performance while also maintaining pricing stability. Investors can tune the bandwidth choice for their desired tradeoff between flexibility and stability.

For fixed-price and hybrid callable bonds, we expand the local linear regression to include the option-induced yield spread as an additional explanatory variable,

$$\mathbf{y}_t = [\mathbf{1}, \mathbf{c}_t, \mathbf{m}_t, \mathbf{v}_t, \mathbf{q}_t, \mathbf{s}_t]\, \mathbf{b}(\mathbf{c}_t, \mathbf{m}_t, \mathbf{v}_t, \mathbf{q}_t, \mathbf{s}_t) + \mathbf{e}_t, \tag{15}$$

where the weighting kernel is also expanded to include the distance in the yield spread,

$$w_{ik}^c = w_{ik}^n \exp\left(-\frac{(\ln(1+s_i) - \ln(1+s_k))^2}{2h_s^2}\right). \tag{16}$$

Similar to volatility and liquidity cost, we winsorize the yield spread estimate at its 99th percentile and build the spread distance with a displaced log transformation $\ln(1+s)$. The transformation stays at zero when the yield spread is zero, but tempers the magnitude when the spread estimate becomes large, thus effectively increasing smoothing at the sparsely populated right tail region. We set the bandwidth h_s on the yield spread distance to three times the default choice.

3.1. Explanatory power

We leave the first 12 months of the sample for constructing the volatility and liquidity cost estimators and start the yield model estimation in July 2003 for a total of 224 months. At each month, we construct a cross-sectional R^2 measure as one minus the ratio of the mean squared pricing error to the cross-sectional variance of the bond yield. Table 4 reports in the first row of entry (under "All") the time-series statistics of the R^2 estimates. The R^2 estimates average at 95.65%, and vary within a narrow range between 91.77% and 97.86%. The time-series standard deviation (SD) estimate is very small at merely 1.24%.

[Table 4 about here.]

To examine the pricing performance variation across different segments of the data sample, we also construct R^2 estimates on different subsamples, each constructed as one minus the ratio of the subsample's mean squared pricing error to its mean squared yield deviation from the full-sample mean. We first divide the sample into the three industry groups. Among them, the industrials sector has the highest average R^2 estimate at 95.75% with the lowest standard deviation at 1.19%. The utilities sector has the lowest average R^2 at 93.99% and the largest standard deviation at 3.09%.

We also divide the sample based on the credit rating into investment grade (IG) and high yield (HY), with ratings of BBB- or higher classified as IG and with ratings of BB+ and lower classified as HY. The HY sample has a higher average R^2 estimate at 96.51% and a lower standard deviation at 1.22%, whereas the IG bonds have an average R^2 of 93.63% and a standard deviation of 1.89%.

Finally, we divide the sample based on the bond duration, with durations five years and lower classified as short and durations higher than five years as long. The average R^2 is higher for the short-duration group at 96.63%, and lower for the long-duration bonds at 92.79%.

Across the different subdivisions, although the average R^2 estimates do show some variations, the variations are all small. The model performs reasonably well across all industry, rating, and duration buckets. The uniformly well performance shows that with the flexible dependence structure, the chosen risk characteristics can adequately explain the cross-sectional yield variation.

3.2. Conditionally varying dependence structures

At each month, we perform the local linear regression repeatedly for each bond to generate the valuation on that bond. Each regression uses all the observations but with different weightings based on the distance of the risk characteristics to that target bond issuance. To understand how the dependence structure varies with the risk characteristics, we perform the local linear regression targeting different percentile values of each risk characteristic at the 10, 25, 50, 75, and 90th percentiles. Since the conditional loading estimate on each risk characteristic depends on the target levels of all other risk characteristics, the high dimensionality makes it difficult to present the results. Nevertheless, we find that the conditional loading estimate on each risk characteristic varies mainly with the level of that risk characteristic. Therefore, in Table 5, we summarize the time-series averages of the conditional loading coefficient estimates by focusing on the variation of each loading coefficient along the risk level of the corresponding risk characteristic while holding the other risk characteristics at their median levels. Each panel reports the time-series averages of the percentile values of one risk characteristic and the loading coefficient estimates on that risk characteristic conditional on the risk characteristic at the different percentile levels, with β_n denoting the loading estimates on non-callable bonds and β_c the estimates on callable bonds.

[Table 5 about here.]

The two sets of conditional loading estimates vary in similar patterns. The average magnitudes of the loading estimates tend to be smaller for the callable bond sample than for the non-callable bond sample. The call provision reduces the yield sensitivities to the risk characteristics.

The average loading estimates on the numeric rating are all strongly positive, reflecting the general observation that bond yields increase strongly with the numeric rating. Furthermore, the average loading estimates become higher as the numeric rating becomes larger, reflecting a convexly increasing relation between the bond yield and the numeric rating.

The average loading estimates on duration are also highly positive, reflecting the positive term risk premium; nevertheless, the average loading estimates are more positive at shorter durations than at longer durations, resulting in a concavely upward sloping average term structure.

The average loading estimates on the yield change volatility are increasingly positive as the volatility estimates become larger, except at the 90th percentile where the average loading estimates become somewhat lower, potentially due to data sparsity and increased smoothing. Therefore, the historical volatility estimates can differentiate bond yields beyond bond rating and duration, especially when the volatility estimates are high.

The average loading estimates on the liquidity cost measure are small, suggesting that once we control the bond risk along the other risk dimensions such as rating, duration, and volatility, liquidity cost only has a small contribution to the yield level variation.

For callable bonds with fixed call prices, we use a binomial tree to estimate the value of the call provision and quantify its value contribution with the option-induced yield spread. If the non-callable component of a callable bond were priced similarly to other non-callable bonds with the same credit risk, duration, and historical volatility levels, this yield spread would fully adjust the callable bond yield to match the yield of other callable bonds, and the loading estimates on the yield spread would average at one. The last panel in Table 5 shows that the average loading estimates on the yield spread are all lower than one and decline with the percentile values of the spread. The lower estimates can partially be driven by the errors-in-variables issue due to the noise in the estimated yield spread. In addition, the likelihood of being called before expiry effectively shortens the bond's duration and reduces the bond's risk sensitivity.

3.3. Predicting future bond returns with the residual yield

Estimating the bond valuation model decomposes the yield of each bond into two components: (i) a fair value component driven by credit, duration, volatility, liquidity, and the call provision for callable bonds, and (ii) a residual yield component driven by possibly short-term supply-demand distortions and other sources of mispricing. With a well-constructed bond valuation model, the residual yield component can represent mispricing opportunities and the starting point for value investing in the corporate bond market.

To examine the model's potential for value investing, we sort the bonds into quintile portfolios at each month based on the residual yield from the valuation model. Table 6 reports the time-series averages of the bond characteristics of the quintile portfolios in panel A. The sorted yield residual increases from an average of -60 basis points at the first quintile to 67 basis points at the fifth quintile. The quintile portfolios have reasonably uniform distributions in terms of the average fair yield, rating, duration, volatility, liquidity, yield spread, and the optionality exposure. The balanced nature of the fair yield and the risk characteristics across the residual quintiles suggests that the yield residual differences are not driven by systematic risk exposures but reflect mostly idiosyncratic variations. More careful inspection shows that bonds at the first and the fifth quintile tend to have higher fair yield, higher numeric rating, higher volatility, and higher liquidity cost estimates. As such, the absolute magnitudes of the pricing errors in yields tend to be larger for bonds with higher risks.

[Table 6 about here.]

Panels B reports the summary statistics of the quintile portfolio's excess return over the next month, including the mean excess return (Mean), the Newey and West (1987) t-statistics on the mean return (NW), the return volatility (Volatility), and the annualized information ratio (IR). The average excess return increases monotonically with the residual yield quintile, from -0.41% at the first quintile to 10.84% at the fifth quintile. The return volatility estimates are stable across the first four quintiles, but become larger at the fifth quintile. The average excess return per unit risk, as captured by the information ratio estimate, increases strongly and monotonically with the residual yield quintile from -0.07 at the first quintile to 1.18 at the fifth quintile.

The investment return from a corporate bond can come from the carry of the yield itself and the capital gain due to the change in the yield. Since the yield residuals are much smaller than the fair yield level, the average carry contributions are strongly positive across all quintiles. As such, for the first quintile portfolio to generate a negative average excess return, the average yield of the portfolio must go up significantly to induce a large capital loss. Panel C of Table 6 reports the average annualized yield change over the next month (Mean) for each quintile portfolio, the Newey and West (1987)-t value (NW) on the mean yield change, and the annualized yield change volatility (Volatility). As expected, when the residual yield is highly negative at the first quintile, the average yield change is highly positive at 1.19% per annum, with strong statistical significance. As the average residual yield becomes less negative in the second and the third quintiles, the average yield change also declines and becomes less positive. When the average residual yield becomes positive in the fourth and fifth quintiles, suggesting that the bond yields are higher than fair as determined by the valuation model, the bond yield declines on average over the next month, leading to negative average yield change estimates.

Inspecting the absolute magnitude of the mean yield changes shows that the yield reversal behavior tends to be stronger in going up than going down. The average yield increase in the first quintile is much stronger than the average yield decline in the fifth quintile. The yield reversal at the fifth quintile also experiences larger volatility, which contributes to the larger volatility estimate for the quintile portfolio's excess return. Still, the combination of the high carry with the yield drop results in a very high average excess return for the fifth quintile, and a high information ratio estimate despite the high return volatility.

In the last column of the table, we construct the high-minus-low (HL) quintile spread portfolio by being long the underpriced bonds in the fifth quintile and short the overpriced bonds in the first quintile. By partially canceling out the systematic risk exposures, the return volatility of the quintile spread portfolio becomes much lower than the volatility of any of the quintile portfolios. As a result, the quintile spread portfolio not only generates a highly positive average excess return, but also with an extremely high t-value at 7.14 and a very high information ratio at 2.39.

Different from duration, credit, or volatility estimates, which can all differentiate bond return risk but cannot differentiate the average excess return per unit risk in a monotonic pattern, the residual yield does not bear any systematic risk exposures, but purely differentiates the mispricing opportunities and strongly predicts future bond returns per unit risk.

4. A corporate bond return factor model

A bond return factor model can serve at least three purposes. First, the factor structure can identify the major sources of co-movements so that the covariance of the return residuals from the model becomes close to a diagonal matrix. Summarizing the co-movements via a parsimonious factor structure reduces dimensionality and increases the robustness of the covariance matrix estimation. Companies in the same industry share similar businesses and respond similarly to macroeconomic shocks. Therefore, as in classic equity return factor models, we propose to use industry dummy variables to capture the stronger co-movements for bond returns within the same industry. Furthermore, our analysis has shown that bonds with similar durations and credit ratings tend to co-move more strongly. To capture the covariance structure along the duration dimension, we choose a set of pivot duration points at $\overline{m} = 1$, 2, 4, 8, and 16 years, and construct loading functions for each bond i to each duration pivot point k via a Gaussian kernel,

$$\omega_{ik}(m) = \exp\left(-\frac{(\ln m_i - \ln \overline{m}_k)^2}{2h_m^2}\right), \tag{17}$$

where the loading is 100% at the target duration and declines smoothly as the duration distance to target increases. Similarly, to capture the covariance structure along the rating dimension, we choose a set of pivot points at $\overline{c} = 3$, 6, 9, 12, 15, and 18, and construct loading functions for each bond i to each rating pivot point $\overline{c}_j$ via a similar Gaussian kernel,

$$\omega_{ij}(c) = \exp\left(-\frac{\left(c_i - \overline{c}_j\right)^2}{2h_c^2}\right). \tag{18}$$

Compared to the commonly used statistical beta estimates on term spread and default spread, which effectively reduce each risk source to a single risk dimension, the loading matrices in (17) and (18) capture the observed high-dimensional structure in each risk source. The construction converts the three structural risk characteristics, i.e., industry, duration, and rating, directly into an exposure matrix, alleviating the errors-in-variable issue in statistical beta estimates.

Second, the return factor structure can capture the cross-sectional variation in the risk magnitudes of the underlying bond return. Data analysis has shown that the bond return risk varies strongly with credit rating and duration, and that the historical bond risk behavior can also differentiate bond yield and future bond return risk beyond those captured by rating and duration.

Based on these observations, we propose to construct a smoothed return volatility estimator for each bond by pooling the historical excess returns on all bonds over the past 12 months and applying weighting based on distances to the target bond in duration, rating, and yield level as in (3) for the pooled yield-change volatility construction. The volatility estimator captures the bond return risk variation across duration, rating, and yield level, while smoothing out the idiosyncratic noises and resolving the lack of history issue for a large proportion of bonds.

Third, the factor structure can identify "priced" risk factors that explain the risk premium variation, the main focus of most academic studies. Our analysis has shown that the optionality exposure for callable bonds represents a distinct risk source that demands higher risk premiums per unit risk. We also include the liquidity cost measure in the return factor model to control for the pricing of bond illiquidity. Most importantly, the yield residual from the bond valuation model strongly predicts future bond returns without taking on systematic risk exposures. We include the yield residual in the return factor model as a mispricing factor to explain the future bond return variation due to the reversal of bond mispricing.

Formally, we construct the following bond return factor model structure,

$$\mathbf{r}_{t+1} = \omega(\mathbf{g}_t, \mathbf{c}_t, \mathbf{m}_t)\phi_{t+1} + \mathbf{v}_t\psi_{t+1} + \mathbf{o}_t\zeta_{t+1} + \mathbf{mq}_t\gamma_{t+1} + \mathbf{me}_t\eta_{t+1} + \varepsilon_{t+1}, \tag{19}$$

where the model explains the cross-sectional variation of the bond excess return over the next month ($\mathbf{r}_{t+1}$) with the following set of risk exposures and risk characteristics:

1. The high-dimensional loading matrix on industry, credit, and duration, $\omega(\mathbf{g}_t, \mathbf{c}_t, \mathbf{m}_t)$, to capture the stronger co-movements within the same industry group, similar credit rating classes, and similar duration buckets. $\mathbf{g}_t$ denotes an $(N_t \times 1)$ vector classifying the N_t bonds at each month into the three industry groups: industrials, financials, and utilities. We first construct

an $(N_t \times 3)$ industry dummy matrix based on the industry classification, and then construct the joint loading matrix $\omega(\mathbf{g}_t, \mathbf{c}_t, \mathbf{m}_t)$ as the kroner product of the industry dummy matrix, the duration loading matrix in (17), and the credit rating loading matrix in (18).

2. The historical return volatility estimator, $\mathbf{v}_t$, to capture the bond return risk variation across rating, duration, and different yield levels.

3. The optionality exposure, $\mathbf{o}_t$, to capture the market pricing of the optionality exposure to interest-rate volatility variations for callable bonds.

4. The liquidity cost estimate multiplied by the bond duration, $\mathbf{mq}_t$, to capture the market pricing of the liquidity risk. The liquidity cost measure $\mathbf{q}_t$ uses the yield change volatility to capture the interest-rate risk. We multiply $\mathbf{q}_t$ by the bond duration to translate the interest-rate risk to bond return variation.

5. The yield residual from the bond valuation model multiplied by the bond duration, $\mathbf{me}_t$, to capture the bond mispricing and its prediction of future bond returns. The multiplication of the bond duration translates the yield reversal movement to the bond return space.

The return factor model in (19) captures the predictive contribution of the residual yield as a bond mispricing factor on the bond excess return while also controlling for the multi-dimensional structural loading exposures to industry, rating, and duration, and accounting for the historical return risk difference and the differential pricing of the optionality exposure and liquidity risk.

4.1. Variance contribution from return risk factors

We start by estimating the return factor model in (19) with ordinary least square cross-sectional regressions at each month. To understand the contribution of each risk factor to the cross-sectional bond excess return variation, we construct a variance contribution (VC) measure for each explanatory variable. From the cross-sectional return forecasting regression $\mathbf{r}_{t+1} = X_t \mathbf{b}_{t+1} + \varepsilon_{t+1}$, where X_t denotes the risk factor matrix and $\mathbf{b}_{t+1}$ the slope coefficient vector, the proportion of uncentered return variance explained by the risk factors X_t can be written as $\mathbf{b}_{t+1}^{\top} \left(X_t^{\top} X_t\right) \mathbf{b}_{t+1} / \left(\mathbf{r}_{t+1}^{\top} \mathbf{r}_{t+1}\right)$. We decompose this total explained variation into variance contributions from each risk factor k as,

$$\mathrm{VC}_{k,t} = [\mathbf{b}_{t+1}]_k \, [(X_t^{\top} X_t)\mathbf{b}_{t+1}]_k / (\mathbf{r}_{t+1}^{\top} \mathbf{r}_{t+1}). \tag{20}$$

Table 7 reports the time-series statistics on the variance contribution estimates, including the sample average (Mean), standard deviation (SD), and the percentile values at 10, 50, and 90th percentiles. The table aggregates the contributions from the high-dimensional loading matrix $\omega(\mathbf{g}_t, \mathbf{c}_t, \mathbf{m}_t)$ into a single percentage. The last row reports the summation of contributions from all risk factors, which represents the gross R^2 of the forecasting return regression.

[Table 7 about here.]

The structural loading matrix captures the stronger co-movements between bonds in the same industry and similar rating and duration buckets. It explains an average of 25.9% of the return variation. The bond return risk also explains a large proportion of the cross-sectional return variation, with the variance contribution averaging at 12.09%. The contributions from the optionality exposure and the liquidity cost are much smaller, both averaging less than 1%. The yield residual from the bond valuation model, which we label as the mispricing factor, explains 1.42% of the return variation on average. Together, the return factor model can explain an average of 40.62% of the cross-sectional bond excess return variation.

The variance contribution estimates from the different risk factors show large time-series variations. The time-series standard deviation estimates tend to be larger than the mean variance contribution estimates for most risk factors. Nevertheless, in aggregate, the return factor model explains the cross-sectional bond return variation well in most days. The gross R^2 estimates have a time-series standard deviation of 19.17%, about half of the mean R^2 estimate. The estimates vary from 16.44% at the 10th percentile to 66.66% at the 90th percentile.

Figure 3 plots the time series variation of the variance contributions from the different risk sources. We apply a 12-month moving average to the monthly estimates to remove the temporal noise and highlight the broad variation pattern. Panel A compares the variance contribution from the structural loading matrix in the solid line with that from the return risk factor in the dash-dotted line. The contributions vary greatly across different market conditions. When the market is experiencing turmoils around 2009, 2016, and 2020, the return risk factor dominates the variance contribution whereas the contributions from the loading matrix become small. During these time periods, the structural characteristics no longer fully reflect the risk behaviors of the bond issuances. The return volatility estimator becomes the more up-to-date risk indicator and the key differentiator of future bond returns.

[Figure 3 about here.]

There are also time periods when the structural loading matrix dominates the variance contribution whereas the contribution from the return risk factor becomes very small. This is the case when the market is calm, and the historical return volatility estimator is not only low, but also less differentiating of the bond's future return variation. The structural loading matrix on industry, duration, and credit rating becomes the most differentiating risk factors.

Panel B plots the time-series variation of the variance contributions from the optionality exposure, liquidity, and the mispricing factor. The average contributions from these three risk sources are much smaller than that from the structural loading matrix and the return risk factor. The optionality exposure contributes positively to the return variation mainly during the first half of the sample. Its contribution becomes much smaller in the second half. The contribution from the liquidity cost factor is not only small, but also unstable, and can swing from positive to negative. By comparison, the residual yield from the valuation model can explain a small but stable proportion of the return variation due to the yield reversal from the bond mispricing. The variance contribution estimates stay positive and steady over the whole sample period.

4.2. The excess return behavior of the bond factor portfolios

A cross-sectional forecasting return regression on the return factor model generates slope coefficient estimates on the risk factors. The slope estimate on each risk factor represents the excess return over the next month on the factor portfolio that targets a unit exposure to the corresponding risk factor while being neutral to all other risk factors. Table 8 summarizes the factor portfolio return behaviors on the left side of the table under panel (i). The statistics include the annualized mean excess return (Mean), the Newey and West (1987) t-statistics on the mean excess return (NW), the annualized return volatility (Volatility), and the annualized information ratio (IR). On the right side of the table under panel (ii), we re-normalize the portfolio weights of each factor portfolio at each month so that the absolute notional investment in each portfolio sums to two dollars. By fixing the notional investment amount, the magnitudes of the mean excess returns become more comparable to the magnitude of the average excess returns on the quintile spread portfolio in Table 6, which invests one dollar long in the fifth quintile and one dollar short in the first quintile. The table omits the statistics on the coefficients of the high-dimensional loading matrix $\omega(\mathbf{g}_t, \mathbf{c}_t, \mathbf{m}_t)$, but focuses on the factor portfolios targeting the return risk, the optionality exposure, the liquidity cost, and the bond mispricing factor.

[Table 8 about here.]

Panel A reports the factor portfolio return behaviors obtained from the equal-weighted ordinary least square regression. The four factor portfolios all generate positive average excess returns, but with different magnitudes and statistical significance. The different normalizations in the two panels lead to different magnitudes for the mean excess returns, but the scale-free information ratio estimates are broadly similar across the two normalizations.

Per dollar notional investment, the return risk portfolio generates an average excess return of 3.82% per annum, but with a return volatility estimate about three times as high at 10.34%, resulting in low statistical significance and low information ratio. The optionality portfolio generates the lowest average excess return at 0.53% per annum per unit notional, with the lowest statistical significance and information ratio. The liquidity cost portfolio generates an average excess return of 3.02%.

Different from these portfolios targeting different risk sources and earning risk premiums, the portfolio on the mispricing factor represents a value-investing statistical arbitrage portfolio that targets the bond mispricing while being neutral to all systematic risk sources. This value investing portfolio generates the highest average excess return at 9.88% per annum per dollar investment, and with the lowest return volatility estimate at 2.52%. The mean excess return estimate is highly statistically significant. The information ratio estimates are also extremely high, at 4.03 when targeting fixed risk exposure and 3.92 when targeting a fixed notional investment.

4.3. Risk-parity weighting in portfolio construction

While the cross-sectional forecasting regression on the return factor model directly generates the excess return on each factor portfolio, we can more generally illustrate the construction of the kth factor portfolio in a classic mean-variance optimization setting,

$$\min_{\mathbf{w_t}} \frac{1}{2} \mathbf{w}_t^\top \Sigma_t \mathbf{w}_t, \quad \text{s.t.} \qquad X_t^\top \mathbf{w}_t = \mathbf{1}_k, \tag{21}$$

where $\mathbf{w}_t$ denotes the $(N_t \times 1)$ bond portfolio weight vector, Σ_t denotes the conditional return covariance matrix, and X_t denotes the $(N_t \times K)$ matrix of the return risk factors as the regressors of the return factor model, and $\mathbf{1}_k$ denotes an $(K \times 1)$ vector of zeros except one for the k element. The mean-variance optimization problem in (21) can be solved analytically as,

$$\mathbf{w_t} = \Sigma_t^{-1} X_t \left(X_t^\top \Sigma_t^{-1} X_t \right)^{-1} \mathbf{1}_k. \tag{22}$$

The equal-weighted ordinary least square regression on the return factor model generates the portfolio weight under the assumption of iid return variance $\Sigma_t = I\sigma^2$. This iid return variance assumption deviates strongly from reality. In particular, our data analysis has shown that the return volatility can increase strongly with bond duration, credit rating, and the yield level. To construct the portfolios more in line with the mean-variance principles, we perform a weighted least square regression estimation of the return factor model, where we scale the excess return of each bond and its regressors by the return volatility estimator on that bond. This volatility scaling amounts to imposing equal risk weighting. It solves the portfolio weight according to the mean-variance construction in (22) by assuming a diagonal covariance matrix $\Sigma_t = \left\langle \mathbf{v}_t^2 \right\rangle$, with each diagonal element being the return variance estimator of the corresponding bond.

Panel B of Table 8 reports the factor portfolio return behaviors obtained from this equal risk weighting estimation. Through the weighting, the portfolio on the return risk factor effectively becomes a risk-parity portfolio while being neutral to other exposures. Indeed, without controlling for other risk factors, the weights on this portfolio will simply be proportional to the reciprocal of the return volatility estimator. As has been found in the literature, e.g., Asness, Frazzini, and Pedersen (2012), forming portfolios with risk parity can enhance diversification and increase the average risk-adjusted excess return. When targeting a unit risk exposure, the annualized information ratio for the return risk portfolio increases strongly from 0.17 with equal dollar weighting to 1.27 with risk-parity weighting.

Equal risk weighting also increases the average excess return and the information ratio on the optionality portfolio, but it reduces the average excess return on the liquidity portfolio to virtually zero, suggesting that the idiosyncratic liquidity risk does not ask for a significant average risk premium when the portfolio is properly adjusted for return risk exposures.

With equal risk weighting, the value-investing portfolio becomes better diversified. The average excess return per unit notional becomes lower at 3.72%, but the return volatility declines even more. The annualized information ratio increases to 4.08 by targeting fixed notional investment and to 4.47 by targeting fixed risk exposure.

4.4. Accounting for same-issuer bond co-movements

The data sample includes 36,102 bonds issued by 3,814 unique issuers. The multiple bond issuances from the same issuer necessarily show stronger co-movements as they are affected by the risk variations of the same issuing company. The return factor model in (19) uses the loading matrix $\omega(\mathbf{g}_t, \mathbf{c}_t, \mathbf{m}_t)$ to control for stronger bond co-movements within the same industry, rating, and duration buckets, but it does not fully account for the stronger co-movements between bonds from the same issuers.

To examine our hypothesis on the stronger co-movements between same-issuer bonds, we construct two average return correlation estimators, one between bonds from the same issuers ($\rho_{SI,t+1}$) and the other between bonds from different issuers ($\rho_{DI,t+1}$),

$$\rho_{SI,t+1} = \frac{\sum_{i\neq j, I_i = I_j} r_{i,t+1} r_{j,t+1}}{\sum_{i\neq j, I_i = I_j} \left| r_{i,t+1} r_{j,t+1} \right|}, \quad \rho_{DI,t+1} = \frac{\sum_{i\neq j, I_i \neq I_j} r_{i,t+1} r_{j,t+1}}{\sum_{i\neq j, I_i \neq I_j} \left| r_{i,t+1} r_{j,t+1} \right|}, \tag{23}$$

where (i, j) are indicators of bond issuances and (I_i, I_j) denote the issuers of the correspond bonds. The two average correlation estimators are constructed at each month on the set of return pairs from the same issuers and different issuers, respectively. Table 9 reports the time-series average (Mean) and standard deviation (SD) of the correlation estimates in the first row of entry, with panel A for the same-issuer return correlation and panel B for the different-issuer correlation estimates. The different-issuer return correlation estimates averages around 35.92%. The same-issuer correlation estimates average nearly twice as high at 61%. Consistent with our hypothesis, the co-movements are much stronger between bonds from the same issuers than from different issuers.

[Table 9 about here.]

Furthermore, to examine how much the return factor model can capture the return co-movement structure, we also construct the two sets of correlation estimators on the return residuals (ε_{t+1}) from the return factor model regression. Table 9 reports the correlation estimates on the return residuals estimated with different weighting schemes. The estimates show that the return factor structure is quite effective in removing correlations between bonds from different issuers. Regardless of the weighting scheme, the residuals from the return factor model have little cross-correlation left between bonds from different issues.

Nevertheless, the return factor structure does not fully account for the same-issuer bond co-movements. The time-series averages of the correlation estimates between the return residuals from the same bond issuers average at 20.12% from the equal dollar weighting estimation and 24.42% from the equal risk weighting estimation.

To account for the stronger same-issuer bond co-movements, we re-estimate the return factor model with a covariance matrix Σ_t that incorporates a 25% correlation between bonds from the same issuers while keeping the correlation to zero between bonds from different issuers. The process is similar to the classic two-stage least square regression. The equal risk weighting estimation can be regarded as the first stage where we use a diagonal covariance matrix $\Sigma_t = \left\langle \mathbf{v}_t^2 \right\rangle$. In the second stage, we set the same-issuer bond correlation to the average estimate on the return residuals from the first stage estimation. Panel C of Table 8 reports the excess return behaviors of the factor portfolios estimated with the newly constructed covariance matrix with a 25% same-issuer correlation. The added same-issuer correlation enhances the diversification of all four factor portfolios. Compared to panel B without the correlation adjustment, the return volatility estimates on all factor portfolios become lower. The information ratio estimates more than double on the return risk factor portfolio and the optionality portfolio. The information ratio on the value-investing portfolio also increases to 6.83 by targeting fixed risk and 4.94 by targeting fixed notional investment.

Table 9 shows that the return residuals from this second-stage estimation generate an average same-issuer correlation estimate of 41.7%, much higher than the 25% input in the covariance construction. One more iteration with the same-issuer correlation input set at 42% generates return residuals with an average same-issuer correlation estimate of 43.69%, very close to the input value. The higher same-issuer correlation input in the covariance construction further reduces the return volatility and enhances the information ratio estimates of the factor portfolios, as shown in panel D of Table 8. The information ratio estimate on the value-investing portfolio increases to 7.23 by targeting fixed risk and 4.98 by targeting fixed notional investment.

5. Concluding remarks

The U.S. corporate bond market has been undergoing rapid transformations from an opaque market with mostly private dealings between broker dealers and long-term investors to one that is increasingly electronic and publicly accessible. It is increasingly becoming the new frontier for both electronic market making and systematic investment.

This paper strives to identify value-based systematic investment opportunities in this new frontier market through the joint construction of a bond valuation model and a return factor model. The valuation model explains the cross-sectional variation of bond yields with a flexible local linear functional form in bond risk characteristics. The return factor model embeds the residual yield from the valuation model as a mispricing factor, while accounting for stronger co-movements between bonds from the same industry, similar rating classes, and similar duration segments, as well as differential market pricing for bond return risk, liquidity cost, and optionality exposure. The slope coefficient on the mispricing factor in the return factor model captures the ex post excess return on the value-investing portfolio that targets a unit exposure to the identified mispricing opportunities while being neutral to all systematic risk exposures.

Historical analysis shows that the bond valuation model explains the cross-sectional bond yield variation very well, with the explanatory power averaging at 95.65% and varying over a narrow range both over time and across different industries, rating classes, and duration. The pricing residuals from the valuation model strongly predict future bond excess returns without bearing systematic risk exposures.

The return factor structure can explain over 40% of the cross-sectional bond excess return variation over the next month. The structural loading matrix we construct on industry, rating, and duration is the major contributor of the explanatory power, especially during calm market conditions. The historical return risk estimator is also a key differentiator of future bond returns, especially during market turmoils. The identified bond mispricing opportunities show strong predictability of future bond returns while only explaining a small proportion of the bond return variation. As a result, the value-investing portfolio targeting the mispricing opportunities while being neutral to systematic risk exposures generates high average excess returns with low return volatility, leading to extremely high risk-adjusted investment performance.

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.

Acharya, V. V, and J. Carpenter. 2002. Corporate Bond Valuation and Hedging with Stochastic Interest Rates and Endogenous Bankruptcy. *Review of Financial Studies* 15(5):1355–1383.

Altman, E. I. 1968. The Prediction of Corporate Bankruptcy: A Discriminant Analysis. *Journal of Finance* 23(1):193–194.

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

Amihud, Y. 2002. Illiquidity and Stock Returns. *Journal of Financial Markets* 5(1):31–56.

Asness, C. S., A. Frazzini, and L. H. Pedersen. 2012. Leverage Aversion and Risk Parity. *Financial Analysts Journal* 68(1):47–59.

Bai, J, and L. Wu. 2016. Anchoring Corporate Credit Default Swap Spreads to Firm Fundamentals. *Journal of Financial and Quantitative Analysis* 51(5):1521–1543.

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

Barberis, N, and M. Huang. 2008. Stocks as Lotteries: The Implications of Probability Weighting for Security Prices. *American Economic Review* 98(5):2066–2100.

Becker, B., M. Campello., V. Thell, and D. Yan. 2024. Credit Risk, Debt Overhang, and the Life Cycle of Callable Bonds. *Review of Finance* forthcoming.

Black, F, and J. C. Cox. 1976. Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. *Journal of Finance* 31(2):351–367.

Brennan, M. J, and E. S. Schwartz. 1977. Convertible Bonds: Valuation and Optimal Strategies for Call and Conversion. *Journal of Finance* 32(5):1699–1715.

Calvet, L. E., A. J. Fisher, and L. Wu. 2018. Staying on Top of the Curve: A Cascade Model of Term Structure Dynamics. *Journal of Financial and Quantitative Analysis* 53(2):937–963.

Cao, J., A. Goyal., X. Xiao, and X. Zhan. 2023. Implied Volatility Changes and Corporate Bond Returns. *Management Science* 69(3):1375–1397.

Carr, P, and L. Wu. 2009. Variance Risk Premiums. *Review of Financial Studies* 22(3):1311–1341.

Carr, P, and L. Wu. 2023. Decomposing Long Bond Returns: A Decentralized Theory. *Review of Finance* 27(3):997–1026.

Chen, L., D. A. Lesmond, and J. Wei. 2007. Corporate Yield Spreads and Bond Liquidity. *Journal of Finance* 62(1):119–149.

Collin-Dufresne, P., R. S. Goldstein, and J. S. Martin. 2001. The Determinants of Credit Spread Changes. *Journal of Finance* 56(6):2177–2207.

Conrad, J., N. Kapadia, and Y. Xing. 2014. Death and Jackpot: Why Do Individual Investors Hold Overpriced Stocks?. *Journal of Financial Economics* 113(3):455–475.

Correia, M., S. Richardson, and İ. Tuna. 2012. Value Investing in Credit Markets. *Review of Accounting Studies* 17:572–609.

Covitz, D, and C. Downing. 2007. Liquidity or Credit Risk? The Determinants of Very Short-term Corporate Yield Spreads. *Journal of Finance* 62(5):2303–2328.

Dai, Q, and K. Singleton. 2000. Specification Analysis of Affine Term Structure Models. *Journal of Finance* 55(5):1943–1978.

Dang, T. D., F. Hollstein, and M. Prokopczuk. 2023. Which Factors for Corporate Bond Returns?. *Review of Asset Pricing Studies* 13(4):615–652.

De Jong, F, and J. Driessen. 2012. Liquidity Risk Premia in Corporate Bond Markets. *Quarterly Journal of Finance* 2(2):1250006.

Dick-Nielsen, J., P. Feldhütter, and D. Lando. 2012. Corporate Bond Liquidity Before and After the Onset of the Subprime Crisis. *Journal of Financial Economics* 103(3):471–492.

Dick-Nielsen, J., P. Feldhütter., L. H. Pedersen, and C. Stolborg. 2023. Corporate Bond Factors: Replication Failures and a New Framework. Working paper, Copenhagen Business School.

Dickerson, A., P. Mueller, and C. Robotti. 2023. Priced Risk in Corporate Bonds. *Journal of Financial Economics* 150(2):103707.

Dickerson, A., C. Robotti, and G. Rossetti. 2024. Common Pitfalls in the Evaluation of Corporate Bond Strategies. Working paper, UNSW Business School and Warwick Business School.

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

Edwards, A. K., L. E. Harris, and M. S. Piwowar. 2007. Corporate bond market transaction costs and transparency. *Journal of Finance* 62(3):1421–1451.

Eom, Y. H., J. Helwege, and J.-Z. Huang. 2004. Structural Models of Corporate Bond Pricing: An Empirical Analysis. *Review of Financial Studies* 17(2):499–544.

Eraker, B, and M. Ready. 2015. Do Investors Overpay for Stocks with Lottery-Like Payoffs? An Examination of the Returns on OTC Stocks. *Journal of Financial Economics* 115(3):486–504.

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.

Gatheral, J, and A. Schied. 2013. Dynamical Models of Market Impact and Algorithms for Order Execution. in *Handbook of Systemic Risk*, ed. by J. P. Fouque, and J. A. Langsam. Cambridge London.

Gebhardt, W. R., S. Hvidkjaer, and B. Swaminathan. 2005. Stock and Bond Market Interaction: Does Momentum Spill Over?. *Journal of Financial Economics* 75(3):651–690.

Geske, R. 1977. The Valuation of Corporate Liabilities as Compound Options. *Journal of Financial and Quantitative Analysis* 12(4):541–552.

Grinold, R. C, and R. N. Kahn. 1999. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk*. McGraw-Hill Education New York 2nd edn.

Guntay, L., N. Prabhala, and H. Unal. 2004. Callable Bonds, Interest-rate Risk, and the Supply Side of Hedging. Working paper, FDIC, Johns Hopkins Carey Business School, and University of Maryland.

Houweling, P., A. Mentink, and T. Vorst. 2005. Comparing Possible Proxies of Corporate Bond Liquidity. *Journal of Banking and Finance* 29(6):1331–1358.

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

Hull, J, and A. White. 1996. Using Hull-White Interest Rate Trees. *Journal of Derivatives* 3(3):26–36.

Israel, R., D. Palhares, and S. Richardson. 2018. Common Factors in Corporate Bond Returns. *Journal of Investment Management* 16(2):17–46.

Jameson, M., T.-H. D. King, and A. Prevost. 2021. Top Management Incentives and Financial Flexibility: The Case of Make-whole Call Provisions. *Journal of Business Finance & Accounting* 48(1–2):374–404.

Jarrow, R., H. Li., S. Liu, and C. Wu. 2010. Reduced-form Valuation of Callable Corporate Bonds: Theory and Evidence. *Journal of Financial Economics* 95(2):227–248.

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

Kargar, M., B. Lester., D. Lindsay., S. Liu., P.-O. Weill, and D. Zúñiga. 2021. Corporate Bond Liquidity During the COVID-19 Crisis. *Review of Financial Studies* 34(11):5352–5401.

Kelly, B., D. Palhares, and S. Pruitt. 2023. Modeling Corporate Bond Returns. *Journal of Finance* 78(4):1967–2008.

Kish, R. J, and M. Livingston. 1993. Estimating the Value of Call options on Corporate Bonds. *Journal of Applied Corporate Finance* 6(3):95–99.

Kumar, A. 2009. Who Gambles in the Stock Market?. *Journal of Finance* 64(4):1889–1933.

Leland, H. E. 1994. Risky Debt, Bond Covenants and Optimal Capital Structure. *Journal of Finance* 49(4):1213–1252.

Lester, B. 2021. When COVID-19 Reached the Corporate Bond Market. Working paper, Federal Reserve Bank of Philadelphia.

Lin, H., J. Wang, and C. Wu. 2011. Liquidity Risk and Expected Corporate Bond Returns. *Journal of Financial Economics* 99(3):628–650.

Litterman, R, and J. Scheinkman. 1991. Common Factors Affecting Bond Returns. *Journal of Fixed Income* 1(1):54–61.

Longstaff, F. A., S. Mithal, and E. Neis. 2005. Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market. *Journal of Finance* 60(5):2213–2253.

Longstaff, F. A, and B. A. Tuckman. 1994. Calling Nonconvertible Debt and the Problem of Related Wealth Transfer Effects. *Financial Management* pp. 21–27.

Ma, L., D. Streitz, and F. Tourre. 2023. The Art of Timing: Managing Sudden Stop Risk in Corporate Credit Markets. Working paper, Halle Institute for Economic Research and Copenhagen Business School.

Mann, S. V, and E. Powers. 2003a. What is the Cost of a Make-whole Call Provision. *Journal of Bond Trading and Management* 1:315–325.

Mann, S. V, and E. A. Powers. 2003b. Indexing a Bond's Call Price: An Analysis of Make-whole Call Provisions. *Journal of Corporate Finance* 9(5):535–554.

Mauer, D. C. 1993. Optimal Bond Call Policies Under Transactions Costs. *Journal of Financial Research* 16(1):23–37.

Merton, R. C. 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance* 29(1):449–470.

Nashikkar, A. J, and M. G. Subrahmanyam. 2007. Latent Liquidity and Corporate Bond Yield Spreads. Nyu working paper no. fin-07-013 New York University.

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

Park, M, and S. P. Clark. 2015. A Reduced-Form Model for Valuing Bonds with Make-Whole Call Provisions. *Applied Mathematical Finance* 22(6):499–521.

Powers, E. 2021. The Optimality of Call Provision Terms. *Management Science* 67(10):6581–6601.

Powers, E, and S. Tsyplakov. 2008. What Is the Cost of Financial Flexibility? Theory and Evidence for Make-Whole Call Provisions. *Financial Management* 37(3):485–512.

Powers, E. A, and S. Sarkar. 2013. Setting the Optimal Make-whole Call Premium. *Applied Financial Economics* 23(6):461–473.

Simonoff, J. S. 1996. *Smoothing Methods in Statistics*. Springer-Verlag New York.

van Binsbergn, J. H., Y. Nozawa, and M. Schwert. 2023. Duration-Based Valuation of Corporate Bonds. *Review of Financial Studies* forthcoming.

Vu, J. D. 1986. An Empirical Investigation of Calls of Non-convertible Bonds. *Journal of Financial Economics* 16(2):235–265.

Wu, L, and Y. Xu. 2023. Separate Risk from Optionality. Working paper, Baruch College.

Xu, Q. 2018. Kicking Maturity Down the Road: Early Refinancing and Maturity Management in the Corporate Bond Market. *Review of Financial Studies* 31(8):3061–3097.

Table 1
Bond yield and return variation across bond duration and credit rating
At each month, we form bond portfolios targeting different duration and credit rating levels. Entries report each portfolio's average yield and the summary statistics of the portfolio's excess return over the next month, including the annualized mean excess return (Mean), the annualized return volatility (Volatility), and the annualized information ratio (IR). The bond yield, the mean excess return, and return volatility are all represented in percentage points. Panel A varies the target duration of the portfolio while holding the credit rating fixed at BBB+. Panel B varies the target rating while holding the target duration fixed at four years.

Risk	Bond yield	Excess return over the next month		
Measures	Mean	Mean	Volatility	IR
A. Duration				
1	2.99	1.88	3.03	0.62
2	3.35	2.85	3.57	0.80
4	4.10	4.34	5.32	0.82
8	4.89	5.26	7.31	0.72
16	5.28	5.25	8.72	0.60
B. Credit rating				
AAA	3.32	3.20	4.14	0.77
AA	3.45	3.36	4.47	0.75
A	3.77	3.85	4.96	0.78
BBB	4.34	4.63	5.56	0.83
BB	5.45	5.67	6.96	0.81
B	7.15	6.91	9.72	0.71
CCC	8.98	8.25	12.89	0.64

Table 2
Bond yield and return variation with yield volatility and liquidity cost
At each month, we form bond portfolios targeting the 10, 25, 50, 75, and 90th percentile values of a statistical risk measure while holding the target duration fixed at four year and the target rating at BBB+. The two panels consider two statistical risk measures: (A) the historical yield-change volatility estimator and (B) the liquidity cost measure. The statistics include the time-series averages of the percentile values of the corresponding risk measure, the average yield of the portfolios, and the summary statistics of the portfolio excess return over the next month, including the annualized mean excess return (Mean), the annualized return volatility (Volatility), and the annualized information ratio (IR).

Percentile	Risk	Bond yield	Excess return over the next month		
	Measures	Mean	Mean	Volatility	IR
	A. Historical volatility				
10	0.85	3.45	2.40	3.90	0.62
25	1.02	3.49	2.54	4.05	0.63
50	1.33	3.69	2.95	4.37	0.67
75	2.16	4.35	4.61	6.55	0.70
90	4.12	5.80	6.61	12.74	0.52
	B. Liquidity cost				
10	0.35	3.57	2.93	4.48	0.65
25	0.61	3.69	3.19	4.31	0.74
50	1.26	3.87	4.05	5.08	0.80
75	3.24	4.23	3.36	5.66	0.59
90	7.00	4.83	4.64	8.95	0.52

Table 3
Quintile portfolios constructed on option value and risk contributions
At each month, for the subsample of callable bonds with fixed strike prices, we form quintile portfolios based on (i) the option-induced yield spread and (ii) the optionality exposure. Panel A reports the time-series averages of the yield spread and the optionality exposure of each quintile portfolio. Panel B reports the time-series averages of the portfolio characteristics, including the average numeric rating, duration, and bond yield. Panels C reports the statistics on the portfolio excess return over the next month, including the annualized mean percentage excess return (Mean), the annualized percentage return volatility (Volatility), and the annualized information ratio (IR).

Contribution	(i) Option-induced yield spread					(ii) Optionality exposure				
Quintile	1	2	3	4	5	1	2	3	4	5
	A. Option value and risk contribution									
Spread	0.44	0.85	1.17	1.55	2.77	1.35	1.16	1.27	1.38	1.62
Exposure	0.49	0.63	0.66	0.68	0.76	0.27	0.49	0.62	0.76	1.08
	B. Bond characteristics									
Rating	8.96	8.99	9.18	10.05	12.22	8.80	9.36	10.07	10.09	11.10
Duration	8.90	8.53	7.41	6.54	5.67	4.71	6.23	7.63	9.88	8.61
Yield	6.87	5.90	5.58	5.72	6.40	5.24	5.51	6.05	6.51	7.17
	C. Excess return over the next month									
Mean	7.60	5.58	5.17	3.18	1.50	-0.26	2.98	4.21	5.95	10.15
Volatility	16.12	12.94	7.79	6.89	6.17	7.55	7.46	10.09	11.59	14.48
IR	0.47	0.43	0.66	0.46	0.24	-0.03	0.40	0.42	0.51	0.70

Table 4
The explanatory power of the bond valuation model
Entries report the time-series statistics on the cross-sectional R^2 estimates from the bond valuation model, constructed as one minus the ratio of the mean squared pricing error to the cross-sectional variance of the bond yield. The statistics include the average (Mean), the standard deviation (SD), the minimum, and the maximum of the time series. In addition to the R^2 estimates on the whole sample ("All"), we also compute the R^2 estimates on different subsamples: (i) across the three industry groups, (ii) the investment grade (IG) versus the high yield (HY) bonds, and (iii) short duration (no longer than five years) versus long duration (greater than five years).

Sample	Mean	SD	Minimum	Maximum
All	95.65	1.24	91.77	97.86
Industrials	95.75	1.19	91.99	98.11
Financials	95.22	1.98	88.39	98.42
Utilities	93.99	3.09	72.76	97.83
IG	93.63	1.89	88.12	96.88
HY	96.51	1.22	91.55	98.40
Short duration	96.63	1.46	91.46	98.81
Long duration	92.79	1.73	87.60	96.99

Table 5
Conditional loading estimates

Entries report the time-series averages of the conditional loading estimates on each risk measure on the yield valuation model, conditional on the target risk measure at the 10, 25, 50, 75, and 90th percentile value, respectively, while holding the other risk measures at their median level. Each panel is for one risk measure. Within each panel, the first row reports the time-series averages of the percentile values of the risk measure. The remaining rows report the time-series averages of the conditional loading estimates, with β_n denoting the loading estimate on the subsample of non-callable bonds, and β_c the loading estimates on the subsample of callable bonds.

Percentile	10	25	50	75	90
Rating	4.56	6.38	8.05	10.48	14.09
β_n	0.13	0.18	0.22	0.29	0.33
β_c	0.12	0.15	0.19	0.22	0.27
Duration	2.03	3.37	5.53	8.96	13.44
β_n	0.38	0.37	0.31	0.20	0.09
β_c	0.14	0.20	0.20	0.15	0.06
Volatility	0.77	0.91	1.16	1.75	3.14
β_n	0.37	0.56	0.80	1.02	0.76
β_c	0.00	0.24	0.64	0.97	0.79
Liquidity cost	0.35	0.61	1.26	3.24	7.00
β_n	-0.02	0.01	0.07	0.11	0.03
β_c	0.04	0.05	-0.00	-0.01	0.00
Yield spread	0.47	0.76	1.17	1.65	2.45
β_c	0.61	0.44	0.39	0.38	0.33

Table 6

Quintile bond portfolios constructed on residual yield

We form quintile bond portfolios based on the residual yield from the bond valuation model. Panel A reports the time-series averages of the bond characteristics of each quintile portfolio. Panels B reports the statistics on the portfolio's excess return over the next month, including the annualized mean excess return (Mean), the Newey-West t-statistics on the average excess return (NW), the annualized return volatility (Volatility), and the annualized information ratio (IR). Panels C reports the statistics on the portfolio's average yield change over the next month, including the annualized mean yield change (Mean), the Newey-West t-statistics on the mean change (NW), and the annualized yield change volatility (Volatility). The last column reports the statistics on the high-minus-low (HL) quintile spread portfolio.

Quintile	1	2	3	4	5	HL
	A. Bond characteristics					
Residual yield	-0.60	-0.19	-0.02	0.14	0.67	1.27
Fair yield	5.30	4.49	4.50	4.69	5.58	0.28
Rating	9.89	8.20	7.77	7.97	9.57	-0.31
Duration	6.41	6.75	6.49	6.69	6.67	0.27
Volatility	2.03	1.43	1.46	1.46	1.97	-0.06
Liquidity cost	2.94	2.49	2.82	2.83	3.12	0.17
Yield spread	0.43	0.42	0.52	0.47	0.38	-0.05
Optionality exposure	0.21	0.19	0.22	0.23	0.21	0.01
	B. Excess return over the next month					
Mean	-0.41	1.76	3.70	6.25	10.84	11.25
NW	-0.27	1.20	2.44	3.64	4.05	7.14
Volatility	5.60	5.66	5.98	6.40	9.19	4.71
IR	-0.07	0.31	0.62	0.98	1.18	2.39
	C. Yield change over the next month					
Mean	1.19	0.58	0.37	-0.13	-0.41	-1.60
NW	2.60	1.63	0.86	-0.32	-0.48	-3.39
Volatility	1.40	1.24	1.36	1.40	3.00	1.84

Table 7

Variance contribution from each return risk factor

At each month, we estimate the return factor model with an ordinary least square regression. Entries report the time-series statistics on the variance contributions from each return risk factor, including the sample average (Mean), standard deviation (SD), and the percentile values at 10, 50, and 90th percentiles. The table aggregates the contributions from the loading matrix on industry, rating, and duration into a single percentage. The last row reports the statistics on the sum of the contributions from all risk factors, which represents the gross R^2 of the regression.

VC, %	Mean	SD	Percentile values		
			10	50	90
Structural Loading	25.90	24.41	0.68	21.17	60.66
Return volatility	12.09	22.50	-8.17	4.51	45.35
Optionality	0.35	2.38	-1.62	-0.01	2.26
Liquidity	0.86	3.75	-2.64	0.07	5.43
Mispricing	1.42	1.56	-0.04	1.08	3.54
R^2	40.62	19.17	16.44	39.87	66.66

Table 8
The excess return behavior of the factor portfolios
Entries report the summary statistics of the excess return series on the factor portfolios, including the annualized mean excess return (Mean), the Newey-West t-statistics on the average (NW), the annualized return volatility (Volatility), and the annualized information ratio (IR). The excess returns on the left side under panel (i) are directly obtained from the return factor model regression targeting a unit exposure to each risk factor. The excess returns on the right side under panel (ii) re-normalize the sum of the absolute portfolio weight of each factor portfolio to to two dollars at each month. Panels A to D represent different weighting schemes for the return factor model estimation.

	(i). Excess return per unit risk				(ii). Excess return per unit notional			
	Mean	NW	Volatility	IR	Mean	NW	Volatility	IR
	A. Equal dollar weighting							
Return risk	0.10	0.64	0.59	0.17	3.82	1.36	10.34	0.37
Optionality	0.11	1.25	0.37	0.29	0.53	0.90	2.64	0.20
Liquidity	0.06	5.07	0.06	0.95	3.02	3.90	3.68	0.82
Mispricing	1.07	10.95	0.27	4.03	9.88	10.78	2.52	3.92
	B. Equal risk weighting							
Return risk	0.80	4.75	0.63	1.27	3.72	3.45	4.24	0.88
Optionality	0.12	1.41	0.37	0.33	0.56	1.27	1.89	0.29
Liquidity	-0.02	-1.88	0.05	-0.34	-0.47	-1.99	1.36	-0.35
Mispricing	1.33	11.85	0.30	4.47	8.29	8.39	2.03	4.08
	C. 25% same-issuer correlation							
Return risk	1.35	9.59	0.46	2.95	5.63	4.36	2.79	2.02
Optionality	0.27	3.45	0.34	0.80	1.10	3.50	1.40	0.79
Liquidity	-0.01	-1.72	0.04	-0.36	-0.24	-1.26	1.02	-0.23
Mispricing	1.62	11.36	0.24	6.83	8.81	7.26	1.78	4.94
	D. 42% same-issuer correlation							
Return risk	1.64	11.03	0.44	3.77	6.01	4.16	2.68	2.24
Optionality	0.32	4.10	0.34	0.95	1.22	3.94	1.34	0.91
Liquidity	-0.01	-1.71	0.04	-0.36	-0.20	-1.12	0.99	-0.20
Mispricing	1.71	11.32	0.24	7.23	8.69	6.98	1.75	4.98

Table 9
Average return correlation between bonds from the same and different issuers
Entries report the time-series average (Mean) and standard deviation (SD) of the return correlation estimates between bonds from the same issuers (panel A) and different issuers (panel B), respectively. The correlations are estimated on both the raw excess returns and the return residuals from the return factor model estimated with different weighting schemes.

Correlation	A. Same issuers		B. Different issuers	
(%)	Mean	SD	Mean	SD
Raw excess return	61.00	26.52	35.92	29.42
Equal dollar weighting	20.12	17.56	-0.31	0.49
Equal risk weighting	24.42	18.77	-0.22	0.55
Same-issuer correlation at 25%	41.70	24.47	0.99	2.86
Same-issuer correlation at 42%	43.69	24.41	1.27	3.24

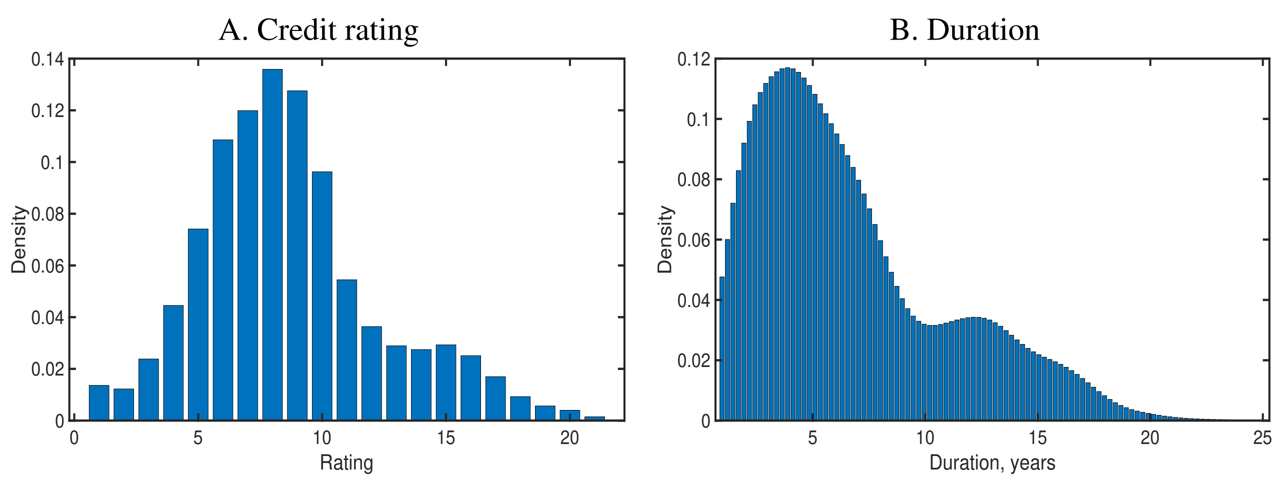

Figure 1

The pooled distribution of bond credit rating and duration.

The pooled distributions of the bond rating in panel A and bond duration in panel B are estimated with a Gaussian kernel.

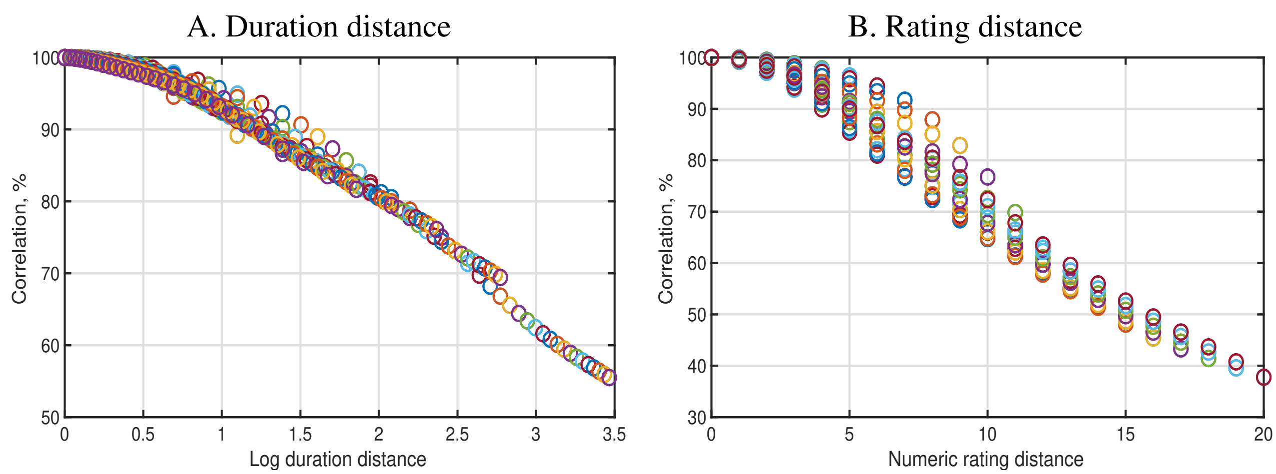

Figure 2
Return correlations decline with log duration and rating distances.
Panel A constructs bond portfolios targeting different durations from one to 32 years at the same fixed rating class of BBB+, and plots the pairwise return correlation estimates as a function of the absolute log duration distance between the two series. Panel B constructs bond portfolios targeting different rating classes from one (AAA) to 21 (C) at the same four-year duration, and plots the pairwise return correlation estimates as a function of the absolute distance in numeric ratings between the two series.

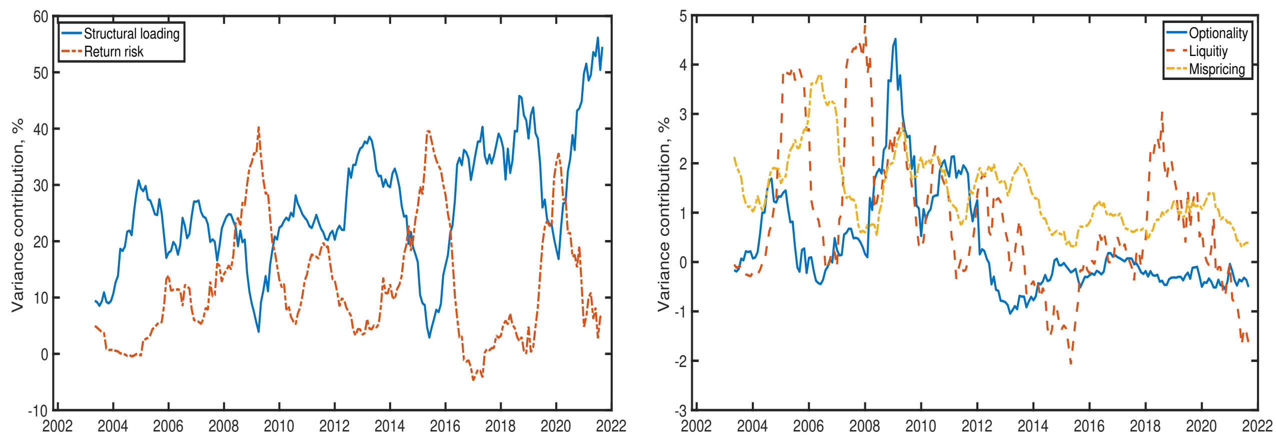

Figure 3

Time-series variation of variance contributions from different risk sources.

Lines plot the 12-month moving average of the time-series variations of the variance contributions from different risk sources. Panel A plots the contributions from the structural loading matrix (solid line) and the return risk factor. Panel B plots the contributions from the optionality exposure (solid line), the liquidity cost (dashed line), and the mispricing factor (dash-dotted line).