RESEARCH

Swiss Finance Institute Research Paper Series N°20-110

Predicting Corporate Bond Returns: Merton Meets Machine Learning

Turan G. Bali
Georgetown University

Amit Goyal
University of Lausanne and Swiss Finance Institute

Dashan Huang
Singapore Management University

Fuwei Jiang
Central University of Finance and Economics (CUFE)

Quan Wen
Georgetown University

Predicting Corporate Bond Returns: Merton Meets Machine Learning*

Turan G. Bali† Amit Goyal‡ Dashan Huang§ Fuwei Jiang¶ Quan Wen‖

Abstract

We investigate the return predictability of corporate bonds using big data and machine learning. We find that machine learning models substantially improve the out-of-sample performance of stock and bond characteristics in predicting future bond returns. We also find a significant improvement in the performance of machine learning models when imposing a theoretically motivated economic structure from the Merton model, compared to the reduced-form approach without restrictions. Overall, our work highlights the importance of explicitly imposing the dependence between expected bond and stock returns via machine learning and Merton model when investigating expected bond returns.

This Version: May 2022

Keywords: Machine learning, big data, corporate bonds, hedge ratio, cross-sectional return predictability
JEL Classification: G10, G11, C13.

*We thank John Y. Campbell, Allan Eberhart, Tom Knox, Jonathan Kluberg, Alejandro Lopez-Lira (our discussant), Christopher Malloy, Markus Pelger, Alberto Rossi, Elvira Sojli, and Derek Vance for their insightful and constructive comments. We also benefited from discussions with seminar participants at the University of Bath School of Management, Arrowstreet Capital, the Center for Financial Markets and Policy and Georgetown University Asset Management Conference, the 2020 Bank of America Global Quant and Innovation Conference, the 2021 Microstructure Exchange seminars, the 2021 BI-SHoF Conference on Asset Pricing and Financial Econometrics, and the 2022 EQD Barcelona conference.

†Robert S. Parker Chair Professor of Finance, McDonough School of Business, Georgetown University, Washington, D.C. 20057. Phone: (202) 687-5388, Fax: (202) 687-4031, Email: Turan.Bali@georgetown.edu

‡Professor of Finance, Faculty of Business and Economics, University of Lausanne and Swiss Finance Institute. Email: amit.goyal@unil.ch

§Associate Professor of Finance, Lee Kong Chian School of Business, Singapore Management University. Email: dashanhuang@smu.edu.sg

¶Professor of Finance, School of Finance, Central University of Finance and Economics. Email: jfuwei@gmail.com

‖Assistant Professor of Finance, McDonough School of Business, Georgetown University, Washington, D.C. 20057. Email: Quan.Wen@georgetown.edu

1 Introduction

A substantial number of stock characteristics have been presented as statistically significant predictors of the cross-section of stock returns since 1970 (Cochrane, 2011). Since then, a few studies show that the majority of the predictive power associated with these characteristics are most likely an artifact of data mining, data snooping, correlated multiple testing, or p-hacking, especially when examined out-of-sample (Harvey, Liu, and Zhu, 2016; Green, Hand, and Zhang, 2017; Linnainmaa and Roberts, 2018; Hou, Xue, and Zhang, 2020). Despite the out-of-sample and post-publication decline of a vast majority of stock characteristics (McLean and Pontiff, 2016), recent studies have shown that machine learning methods are able to generate robust forecasting power to predict stock returns, address the data-snooping concerns, and identify the marginal contribution of new factors relative to the large set of existing ones (Feng, Giglio, and Xiu, 2020; Gu, Kelly, and Xiu, 2020; Kozak, Nagel, and Santosh, 2020; Giglio, Liao, and Xiu, 2021).

Despite the proliferation of stock characteristics or factors to explain the cross-section of stock returns, however, far fewer studies are devoted to predict future returns on corporate bonds. Recent studies examine a few corporate bond characteristics related to default and term betas (Fama and French, 1993; Gebhardt, Hvidkjaer, and Swaminathan, 2005), liquidity risk (Lin, Wang, and Wu, 2011), bond momentum (Jostova, Nikolova, Philipov, and Stahel, 2013), downside risk (Bai, Bali, and Wen, 2019), and long-term reversal (Bali, Subrahmanyam, and Wen, 2021a), which exhibit significant explanatory power for future bond returns. Using standard asset pricing tests such as the OLS cross-sectional regressions, other papers investigate whether well-known equity market anomalies impact the cross-section of corporate bond returns and find mixed evidence on the predictability (Chordia, Goyal, Nozawa, Subrahmanyam, and Tong, 2017; Choi and Kim, 2018).

One common element in most of these studies is that they use standard linear methods to analyze return predictability. However, bondholders are more sensitive to downside risk compared to stockholders (Hong and Sraer, 2013; Bai, Bali, and Wen, 2019). Because of the nonlinear payoffs of corporate bonds and the high correlation between many of the stock and bond characteristics, machine learning is well suited for such challenging prediction problems by reducing the degrees of freedom and condensing redundant variation among a large set of predictors, with an emphasis on variable selection and dimension reduction techniques (Gu, Kelly, and Xiu, 2020).[1]

[1]Recent studies use machine learning techniques to extract information from both the cross-section and time-series of stock returns in identifying the most relevant stock characteristics or factors. For example, Feng, Giglio, and Xiu (2020) propose a model selection method to systematically evaluate the contribution to asset pricing of any new factor, above and beyond what a high-dimensional set of existing factors explains. Lettau and Pelger (2020) develop a risk premium PCA estimator that adds to the traditional PCA objective function a no-arbitrage penalty term that helps price the cross-section of equity returns. Freyberger, Neuhierl, and Weber (2020) introduce a nonparametric method (i.e., the adaptive group LASSO) to study which characteristics provide incremental information for the cross-section of stock returns. Nagel (2021) provides a comprehensive overview of machine learning models and discusses the application of these techniques in empirical research in asset pricing.

In this paper, we provide a comprehensive study on the cross-sectional predictability of corporate bond returns using a large set of stock and bond characteristics. Previous studies, in general, rely on the reduced-form approach that examines cross-sectional bond return predictability, without explicitly linking the functional forms of bond and stock expected returns. In this article, we highlight the importance of imposing a theoretically motivated economic structure when investigating expected bond return predictability, which tends to be largely understudied in the aforementioned research. There are a few reasons to impose an economic structure and investigate the dependence between expected bond and stock returns in a unified framework. First, stocks and bonds issued by the same firm represent claims on the same underlying assets of the firm. Hence, relevant information about the firm should have an impact on both the firm's outstanding stocks and its outstanding bonds, leading to co-movement between individual stock and bond prices. Hence, it is not surprising that their returns should be correlated.[2] Second, the typical workhorse model to analyze the stock-bond connection is Merton (1974) structural credit risk model, which explains how bonds and stocks should be jointly priced. Based on the model of Merton (1974), if a variable/characteristic explains stock returns, then the model places restrictions on the predictability of bond returns from this variable. As a result, motivated by the Merton (1974) model, we impose the dependence between expected returns of bonds and stocks, and compare the forecasted bond returns with such restrictions to the ones obtained from the reduced-form approach that neglects any form of economic structure.

In light of the machine learning methods, we seek to answer the following questions: First, without imposing any economic structure from the Merton (1974) model, do corporate bond characteristics and stock characteristics, individually or combined, predict future bond returns? Do stock characteristics improve the performance of bond-level characteristics in predicting future bond returns? Second, is there any significant improvement in the performance of the machine learning models when imposing the economic structure from the Merton (1974) model, compared to the ones without restrictions? Overall, our results highlight that it is important to explicitly impose the dependence between expected bond and stock return via the Merton (1974) model as such economic structure significantly improves future bond return forecasts. Our results also show that once we impose the Merton (1974) model structure, equity characteristics provide significant improvement above and beyond bond characteristics for future bond returns, whereas the incremental power of equity characteristics for predicting bond returns are quite limited in the reduced-form approach when such economic structure is not imposed.

We first build a comprehensive data library of 43 corporate bond-level characteristics that are motivated by the existing literature on the cross-section of corporate bonds. This list of a broad set

[2] Kwan (1996) indeed finds that stock returns and bond yield changes are positively correlated. Kelly, Palhares, and Pruitt (2022) find that the systematic components of bond and equity returns are roughly twice as integrated as their total returns, whereas idiosyncratic bond and stock returns are substantially less integrated than their systematic counterparts.

of corporate bond return predictors is designed to be representative of (i) bond-level characteristics such as issuance size, credit rating, time-to-maturity, and duration, (ii) proxies of risk such as bond systematic risk, downside risk, and credit risk, (iii) proxies of bond-level illiquidity constructed using daily and intraday transaction data and liquidity risk, (iv) past bond return characteristics such as bond momentum, short-term and long-term reversals, and (v) the distributional characteristics such as return volatility, skewness, and kurtosis.

We then combine them with the 94 stock characteristics used by Green, Hand, and Zhang (2017) and Gu, Kelly, and Xiu (2020). Our final sample of the 137 stock- and bond-level characteristics cover both the equity and debt markets, thus provide a wide range of predictors for corporate bond returns. Focusing on a variety of machine learning methods proposed by Gu, Kelly, and Xiu (2020), we compare and evaluate the out-of-sample performance of alternative machine learning models in predicting the cross-sectional dispersion in future bond returns. The machine learning methods include the dimension reduction models (PCA and PLS), penalized methods (Lasso, Ridge, and Elastic Net), regression trees (Random Forests), and neural networks including the feed forward neural networks (FFN). In addition to these methods, we use the long short-term memory neural network (LSTM) proposed by Hochreiter and Schmidhuber (1997) to capture a long memory effect (Lo, 1991). Moreover, we rely on the forecast combination method (Combination) which averages individual expected return forecasts from the aforementioned sophisticated machine learning models (Rapach, Strauss, and Zhou, 2010; Chen, Pelger, and Zhu, 2019).

We first show that the traditional unconstrained linear regression models such as the OLS fail to deliver statistically significant out-of-sample forecasting power for future corporate bond returns. The standard OLS regression methodology with all 43 bond characteristics produces a negative out-of-sample R-squared (R^2_{OS}), whereas the machine learning models substantially improve the predictive power with R^2_{OS} ranging from 1.85% to 2.37%. Using the Diebold and Mariano (1995) test for differences in out-of-sample predictive accuracy between two models, we find that all machine learning models perform equally well and they significantly outperform the unconstrained OLS model.

To further investigate the economic significance of machine learning approaches, we form corporate bond portfolios based on machine learning forecasts using the 43 bond characteristics. The machine learning driven bond portfolios are constructed based on the one-month-ahead out-of-sample forecasts of bond returns, where the arbitrage (high-minus-low) portfolio corresponds to the long-short portfolio that buys bonds with the highest one-month-ahead expected returns (decile 10) and sells bonds with the lowest one-month-ahead expected returns (decile 1). We find that all machine learning forecasts generate economically and statistically significant return spreads on the arbitrage portfolios, ranging from 0.33% to 0.79% per month, compared to the unconstrained OLS model which delivers the smallest monthly return spread of 0.16%.

We proceed to identify corporate bond characteristics that are important determinants of the cross-section of bond returns, while simultaneously controlling for the many other predictors. Following the ranking and variable importance approach of Kelly, Pruitt, and Su (2019) and Gu, Kelly, and Xiu (2020), we discover influential covariates by measuring the reduction in panel predictive regression R^2_{OS}, while holding the remaining model estimates fixed. This approach allows us to investigate the relative importance of individual bond characteristics for the out-of-sample forecasting performance of each machine learning model. Our results demonstrate that all machine learning models are in close agreement on the most influential bond-level characteristics, which can be classified into four broad categories (i) bond characteristics related to interest rate risk such as duration and time-to-maturity, (ii) risk measures such as downside risk proxied by Value-at-Risk (VaR) and expected shortfall (ES), total return volatility (VOL), and systematic risk proxied by the bond market beta, default beta, and term beta, (iii) bond-level illiquidity measures such as the average bid and ask price (AvgBidAsk), and Amihud and Roll's measures of illiquidity, and (iv) past return characteristics related to bond momentum, short-term reversal, and long-term reversal. To find out which one of the four groups of bond return predictors is the most important determinant of the expected bond returns, we compute the sum of the importance measure of each return predictor for each method, within each characteristic group. We find that the top two most important groups are the characteristics related to bond-level illiquidity and illiquidity risk (i.e., Group III) and risk measures such as downside risk and systematic risk proxies (i.e., Group II).

Then, we examine whether a large number of stock characteristics improve the cross-sectional return predictability of corporate bonds, using the reduced-form approach without explicitly linking the functional forms of bond and stock expected returns via the Merton (1974) model. Recent studies often draw from the well of cross-sectional predictors on a few stock characteristics and find mixed evidence of predictability for corporate bonds (Chordia et al., 2017; Choi and Kim, 2018). Compared to these studies, we extend the candidates to a much larger set of stock characteristics and more importantly, we rely on machine learning methods to reduce redundant variation among predictors that address overfitting bias. We show that all machine learning models substantially improve the forecasting power of stock characteristics for future bond returns compared to the standard OLS, for all sample of bonds.[3] However, the marginal improvement of the forecasting power of stock characteristics relative to bond characteristics is economically small and insignificant, as most machine learning forecasts fail to deliver statistically significant positive return spreads on the long-short bond portfolios.

It is important to note that so far we have only used different machine learning approaches to model bond expected returns, which is a reduced-form approach that does not explicitly link the functional forms of bond and stock expected return. Motivated by Merton (1974) model, we next impose the dependence between expected bond and stock return using hedge ratios. When

[3] The machine learning models using stock characteristics deliver an R^2_{OS} in the range of 1.61% and 2.02%, which is similar to the R^2_{OS} obtained from using bond characteristics, which ranges from 1.85% to 2.37%.

we use regression-based hedge ratios,[4] we find that the machine learning model with such economic structure generates economically and statistically significant return spreads on the long-short bond portfolios, in the range of 0.55% and 0.92% per month, compared to the unconstrained OLS model which delivers the smallest return spread of 0.18%. More importantly, there is significant improvement in the performance of the machine learning models with imposing restrictions, compared to the bond return forecasts obtained without restrictions using bond characteristics alone or the combined stock and bond characteristics.

Finally, we further investigate the predictability of bond returns using Merton (1974) model with hedge ratios estimated with machine learning models. Specifically, we model hedge ratio as a function of bond characteristics and investigate the performance of expected bond return forecasts with Merton (1974) restrictions *and* machine learning estimated hedge ratio. Our results show a positive and statistically significant Diebold-Mariano test statistics for all the machine learning models, compared to the bond return forecasts using only bond characteristics, the combined stock and bond characteristics, or those generated using the Merton model restriction with an exogenously specified hedge ratio. However, the results show that the economic significance of using machine learning estimated hedge ratio is similar to that using exogenously specified hedge ratio, as the return spreads generated from both approaches are similar in economic magnitude and they are not statistically different from each other. Overall, we conclude that it is important to impose Merton model restrictions along the lines of Schaefer and Strebulaev (2008) when estimating bond expected returns, which significantly improves bond return predictability compared to the reduced-form approach that does not explicitly model the dependence between bond and stock expected returns.

The rest of the paper proceeds as follows. Section 2 provides our theoretical motivation, presents the corresponding prediction framework, and describes the performance metrics used to assess the predictive power of stock and bond characteristics. Section 3 describes the data and variables used in our empirical analyses. Section 4 relies on a reduced-form approach that does not explicitly link the functional forms of bond and stock expected returns and investigates the performance of machine learning models in predicting future bond returns without hedge ratios. Section 5 imposes the dependence between expected bond and stock returns via Merton (1974) model and examines the performance of machine learning models in predicting future bond returns using regression-based hedge ratios. Section 6 presents results from predicting future bond returns with machine learning based dynamic hedge ratios. We conclude in Section 7.

[4]Schaefer and Strebulaev (2008) is the first paper to provide a comprehensive investigation of the magnitude and statistical significance of the hedge ratio. Choi and Kim (2018) follow Schaefer and Strebulaev (2008) in terms of the estimation methodology but rely on a different method to estimate the hedge ratio for each firm and for each month based on a rolling regression using monthly returns over the past 36 months.

2 Methodology

2.1 Theoretical Motivation

We present a simple structural model to guide our empirical work. While the typical workhorse model to analyze the stock-bond connection is Merton (1974) structural credit risk model, we follow its extension in Du, Elkamhi, and Ericsson (2019). We assume that the value of the assets of the firm, V_t, is governed by the following stochastic processes:

$$\begin{aligned} dV_t &= rV_t dt + \sigma_t V_t dW_t \\ d\sigma_t^2 &= \kappa(\theta - \sigma_t^2)dt + \gamma\sigma_t dZ_t, \end{aligned} \tag{1}$$

where the initial value of the assets $V_0 > 0$ and r is the risk-free rate. The processes $\{W_t\}$ and $\{Z_t\}$ are two standard Brownian motions under the risk-neutral martingale measure Q and their instantaneous correlation is ρ. κ is the speed of mean reversion, θ is the long-run mean variance, and γ is the volatility parameter for asset variance. The firm issues a single class of debt, a zero-coupon bond, with a face value B payable at time T. Default may happen only at time T, and if default happens, creditors take over the firm without incurring any distress costs and realize an amount V_T. Otherwise, they receive B. Equation (1) differs from Merton (1974) by generalizing the variance of the assets to follow a stochastic process (instead of assuming constant asset variance). Du, Elkamhi, and Ericsson (2019) show that this relaxation can better describe the average credit spreads levels.

When the asset variance is constant, Merton (1974) shows that the creditors take a short position in a put option written on the assets of the borrowing firm with a strike B, the face value of the debt, while the equity holders, who own the firm, borrow the amount B at time 0, and own a put option on the assets of the firm with strike B, equivalently hold a call option on the assets of the firm with strike B. As such, the equity and bond prices at any time t can be explicitly solved by the Black and Scholes (1973) formula.

When the asset variance is stochastic, equity and bond prices cannot be expressed in closed form. In this case, Hull and White (1987) propose an approximation method, which delivers a closed form solution. We follow these authors and approximate the equity and bond prices as

$$E_t = V_t N(d_1) - Be^{-r(T-t)}N(d_2) - \frac{\sqrt{\theta}\gamma}{8\kappa} \cdot \eta_t, \tag{2}$$

where $N(d_1)$ and $\phi(d_1)$ are the standard normal distribution and density functions, respectively. The closed form η_t is provided in equation (A.8) of Appendix A. The debt value is then given by $D_t = V_t - E_t$. We note that the equity price in equation (2) differs from Hull and White (1987), in that our variance follows a Cox, Ingersoll, and Ross (1985) process, while it follows a geometric

Brownian motion in Hull and White (1987).

With equation (2), we can analytically calculate the hedge ratio, following the definition of Schaefer and Strebulaev (2008), as

$$\begin{aligned} h_t &\stackrel{\text{def}}{=} \frac{\partial D_t/\partial V_t}{\partial E_t/\partial V_t} \times \frac{E_t}{D_t} \\ &= \frac{1 - N(d_1) + \gamma^2 \zeta_t}{N(d_1) - \gamma^2 \zeta_t} \frac{E_t}{D_t}. \end{aligned} \tag{3}$$

At the same time, the equity and bond returns have the following relationship:

$$\frac{dD_t}{D_t} - h_t \frac{dE_t}{E_t} = \alpha_t dt. \tag{4}$$

The expressions for ζ_t and α_t are given in equation (A.15) of Appendix A.

Clearly, when the variance of the firm value is constant, i.e., $\gamma = 0$, E_t and h_t reduce to the case in Schaefer and Strebulaev (2008), and $\alpha_t = 0$. Because the bond and equity prices are driven by the firm value V_t only, the two markets are fully integrated or the systemic risk of the bond can be perfectly hedged by the equity. In contrast, when the variance is stochastic, the bond and equity prices are jointly driven by V_t and σ_t^2. The two markets are not fully integrated any more, which is supported by the empirical fact that $\alpha_t \neq 0$.

Equation (4) shows that any prediction of bond returns involves three components: (i) predicting the hedge ratio, (ii) predicting the stock return, and (iii) predicting the 'residual' bond return. This equation forms the basis of our empirical work.

2.2 Prediction Framework

We index assets (either a corporate bond or a stock) by $i = 1, \ldots, N$ and months by $t = 1, \ldots, T$. By definition, the excess return of asset i at time $t+1$, R_{it+1}, is equal to the sum of expected return plus the error term:

$$R_{it+1} = E_t(R_{it+1}) + e_{it+1}, \tag{5}$$

where $E_t(R_{it+1})$ is the time-t expected return. Specifically, let RB and RS denote the realized bond and stock return, respectively. We have:

$$RB_{it+1} = E_t[RB_{it+1}] + eB_{it+1} \tag{6}$$

$$RS_{it+1} = E_t[RS_{it+1}] + eS_{it+1}, \tag{7}$$

where eB and eS are the unexpected bond return and stock return, respectively. Using equation (4), we have

$$E_t(RB_{it+1}) \quad = \quad h_{it} \times E_t(RS_{it+1}) + \alpha_{it}. \tag{8}$$

Substituting equation (8) into equation (6), we get:

$$\begin{aligned} RB_{it+1} &\equiv E_t(RB_{it+1}) + eB_{it+1} \\ &= h_{it} \times E_t(RS_{it+1}) + \alpha_{it} + eB_{it+1} \\ &= h_{it} \times RS_{it+1} + \alpha_{it} + (eB_{it+1} - h_{it} \times eS_{it+1}). \end{aligned} \tag{9}$$

Define $RBmRS_{it+1}$ as the difference between realized bond return (RB) and the product of the hedge ratio and realized stock return ($h \times RS$):

$$\begin{aligned} RBmRS_{it+1} &\stackrel{\text{def}}{=} RB_{it+1} - h_{it} \times RS_{it+1} \\ &= \alpha_{it} + (eB_{it+1} - h_{it} \times eS_{it+1}). \end{aligned} \tag{10}$$

Taking expectation, we see that $E_t(RBmRS_{it+1}) = \alpha_{it}$. We can, thus, express expected bond returns as:

$$E_t(RB_{it+1}) \quad = \quad h_{it} \times E_t(RS_{it+1}) + E_t(RBmRS_{it+1}). \tag{11}$$

The expectations in equation (11) are specified to be flexible functions of characteristics. For instance, a generic time-t expected return, $E_t(R_{it+1})$, is specified to be $E_t(R_{it+1}) = \phi(X_{it})$, where $\phi(\cdot)$ is a flexible function of asset i's P-dimensional characteristics, i.e., $X_{it} = (X_{i1t}, \ldots, X_{iPt})'$. We discuss specific functional forms in the next Section 2.3.

We consider three variations of predicting bond returns:

1. Without the hedge ratios: The benchmark prediction model does not rely on the theoretical framework as outlined in Section 2.1 and implicitly sets the hedge ratio, h_{it}, to zero. As is evident from equations (10) and (11), in this case $RBmRS \equiv RB$. Thus, the prediction task simplifies to predicting just the bond returns with no cross-asset restrictions of the form (8). We specify:

$$E_t(RB_{it+1}) = f_1(X_{it}), \tag{12}$$

where the characteristics X include combinations of bond characteristics, XB, and stock characteristics, XS. Note that even though we do not formally use the hedge ratios in this approach, the stock and the bond market are not assumed to be disconnected. For instance, when X includes both bond and stock characteristics, we allow stock characteristics (predictors of stocks returns) to predict bond returns too. Therefore, this approach can be considered as a reduced-form Merton (1974) approach. We discuss results from this approach in Section 4.

2. With regression-based hedge ratios: In this prediction method, we estimate hedge ratios via regressions of bond returns on stock returns. We then separately estimate $E_t(RS_{it+1}) = \psi_1(X_{it})$ and $E_t(RBmRS_{it+1}) = \psi_2(X_{it})$ and then combine these predictions to obtain the expected bond return as:

$$E_t(RB_{it+1}) = f_2(X_{it}) = \hat{h}_{it} \times \psi_1(X_{it}) + \psi_2(X_{it}). \tag{13}$$

We investigate bond return prediction following this approach in Section 5.

3. With machine learning-based hedge ratios: In this prediction method, we let the hedge ratio itself be a function of characteristics. Thus, we separately estimate three different machine learning models $E_t(RS_{it+1}) = \phi_1(X_{it})$, $E_t(RBmRS_{it+1}) = \phi_2(X_{it})$, and $h_{it} = \phi_3(X_{it})$, and then combine these predictions to obtain the expected bond return as:

$$E_t(RB_{it+1}) = f_3(X_{it}) = \phi_3(X_{it}) \times \phi_1(X_{it}) + \phi_2(X_{it}). \tag{14}$$

Note that the prediction of $E_t(RBmRS_{it+1})$ in this third variant is different from the corresponding prediction in the second variant ($\phi_2(X_{it}) \neq \psi_2(X_{it})$) even if the same set of characteristics is used in both predictions. The reason is that $RBmRS$ in equation (10) is defined using hedge ratio, h_{it}, which is calculated differently in the two approaches. Section 6 provides further details on computations and the associated results.

2.3 Machine Learning and Performance Evaluation

Following Gu, Kelly, and Xiu (2020), we compare and evaluate a variety of machine learning methods, including the ordinary least squares (OLS) with all covariates; penalized linear regression methods such as LASSO, ridge regression (Ridge), and elastic net (ENet); dimension reduction techniques such as principal component analysis (PCA) and partial least square (PLS); random forests (RF); and feed-forward neural network (FFN). In addition to these methods, we use a long short-term memory neural network (LSTM) to capture a long memory effect (Lo, 1991; Hochreiter and Schmidhuber, 1997). Moreover, we rely on the forecast combination method (Combination) which averages individual expected return forecasts from the aforementioned eight machine learning models (Rapach, Strauss, and Zhou, 2010; Chen, Pelger, and Zhu, 2019). We provide a detailed description of these methods in Section OA1 of the Online Appendix.

Following Gu, Kelly, and Xiu (2020), we use the out-of-sample R-squared as the performance metric to assess the predictive power of individual bond return predictors,

$$R^2_{OS} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3} (r_{it+1} - \hat{r}_{it+1})^2}{\sum_{(i,t)\in\mathcal{T}_3} r^2_{it+1}}. \tag{15}$$

The R^2_{OS} statistic pools prediction errors across bonds and over time into a grand panel-level assessment of each model, and it measures the proportional reduction in mean squared forecast error (MSFE) for each model relative to a naive forecast of zero benchmark, which assumes that the one-month-ahead expected return on corporate bonds equals the time $t+1$ risk-free rate. To estimate the out-of-sample R^2_{OS}, we follow the most commonly used approach in the literature and divide our full sample (July 2002 to December 2017) into three disjoint time periods; (i) the first three years of "training" or "estimation" period, $\mathcal{T}_1$, (ii) the second two years of "validation" for tuning the hyperparameters, $\mathcal{T}_2$, and (iii) the rest of the sample as the "test" period, $\mathcal{T}_3$, to evaluate a model's predictive power, which represents the truly out-of-sample evaluation of the model's performance.

We use the mean squared forecast error (MSFE)-adjusted statistic of Clark and West (2007) to test the statistical significance of R^2_{OS}. Considering the potentially strong cross-sectional dependence among individual excess bond returns, we employ the modified MSFE-adjusted statistic based on the cross-sectional average of prediction errors from each model instead of prediction errors among individual returns. The p-value from the MSFE-adjusted statistic tests the null hypothesis that the MSFE of a naive forecast of zero is less than or equal to the MSFE of a machine learning model against the one-sided (upper-tail) alternative hypothesis that the MSFE of a naive forecast of zero is greater than the MSFE of a machine learning model (H_0: $R^2_{OS} \leq 0$ against H_A: $R^2_{OS} > 0$).

To compare the out-of-sample predictive power of two methods, we use the modified Diebold and Mariano (1995) test, which accounts for the potentially strong cross-sectional dependence among individual returns. Specifically, to compare the predictive powers of methods (1) and (2), we define the modified Diebold-Mariano statistic as

$$\mathrm{DM}_{12} = \bar{d}_{12}/\hat{\sigma}_{\bar{d}}, \tag{16}$$

where $\bar{d}_{12}$ and $\hat{\sigma}_{\bar{d}}$ are, respectively, the time-series mean and Newey-West standard error of $d_{12,t+1}$ over the testing sample. $d_{12,t+1}$ is the forecast error differential between the two methods, calculated as the cross-sectional average of forecast error differentials from each model over each period $t+1$,

$$d_{12,t+1} = \frac{1}{n_{3,t+1}} \sum_{i=1}^{n_3} \left(\left(\hat{e}^{(1)}_{it+1}\right)^2 - \left(\hat{e}^{(2)}_{it+1}\right)^2 \right), \tag{17}$$

where $\hat{e}^{(1)}_{it+1}$ and $\hat{e}^{(2)}_{it+1}$ are the return forecast errors for individual asset i at time $t+1$ generated by two methods, and $n_{3,t+1}$ is the number of assets in the testing sample.

3 Data and Variable Definitions

This section first describes the data and key variables used in our empirical analyses and then provides summary statistics for the large set of corporate bond characteristics we construct. Following Bessembinder, Maxwell, and Venkataraman (2006), who highlight the importance of using TRACE transaction data, we rely on the transaction records reported in the enhanced version of TRACE for the sample period from July 2002 to December 2017. The TRACE dataset offers the best-quality corporate bond transactions, with intraday observations on price, trading volume, and buy and sell indicators.[5]

For TRACE data, we adopt the filtering criteria proposed by Bai, Bali, and Wen (2019). Specifically, we remove bonds that (i) are not listed or traded in the US public market; (ii) are structured notes, mortgage backed/asset backed/agency backed/equity-linked; (iii) are convertible; (iv) trade under \$5; (v) have floating coupon rates; and (vi) have less than one year to maturity. For intraday data, we also eliminate bond transactions that (vii) are labeled as when-issued or locked-in or have special sales conditions, (viii) are canceled, (ix) have more than a two-day settlement, and (x) have a trading volume smaller than \$10,000. We then merge corporate bond pricing data with the Mergent FISD to obtain bond characteristics such as the offering amount, offering date, maturity date, coupon rate, coupon type, interest payment frequency, bond type, bond rating, bond option features, and issuer information.

3.1 Corporate Bond Return

The monthly corporate bond return at time t is computed as

$$r_{it} = \frac{P_{it} + AI_{it} + C_{it}}{P_{it-1} + AI_{it-1}} - 1, \tag{18}$$

where P_{it} is the transaction price, AI_{it} is accrued interest, and C_{it} is the coupon payment, if any, of bond i in month t. We denote R_{it} as bond i's excess return, $R_{it} = r_{it} - r_{ft}$, where r_{ft} is the risk-free rate proxied by the one-month Treasury bill rate.

With the TRACE intraday data, we first calculate the daily clean price as the trading volume-weighted average of intraday prices to minimize the effect of bid-ask spreads in prices, following Bessembinder, Kahle, Maxwell, and Xu (2009). We then convert the bond prices from daily to monthly frequency following Bai, Bali, and Wen (2019), who discuss the conversion methods in

[5] We use enhanced TRACE instead of the standard TRACE since it contains uncapped transaction volumes and information on whether the trade is a buy, a sell, or an interdealer transaction, in addition to the information contained in standard TRACE. The improvement of enhanced TRACE over standard TRACE thus allows us to construct a variety measures of bond liquidity using daily and intraday transaction data.

detail. Specifically, our method identifies two scenarios for a return to be realized at the end of month t: (i) from the end of month $t-1$ to the end of month t, and (ii) from the beginning of month t to the end of month t. We calculate monthly returns for both scenarios, where the end (beginning) of the month refers to the last (first) five trading days within each month. If there are multiple trading records in the five-day window, the one closest to the last trading day of the month is selected. If a monthly return can be realized in more than one scenario, the realized return in the first scenario (from month-end $t-1$ to month-end t) is selected.

Corporate bonds occasionally default prior to reaching maturity. If default returns are simply treated as missing observations, return estimates can be overstated, particularly for high-yield bonds and long-term losers. To address this potential return bias, we follow Cici, Gibson, and Moussawi (2017) and Bali, Subrahmanyam, and Wen (2021a) and compute a composite default return for all defaulted bonds. Specifically, we search for any price information on defaulted issues after the default event. We then compute median returns on these defaulted issues in the $(-1, +1)$ month window around the default date and use the median return of -40.17% for defaulting investment-grade (IG) issues and -17.67% for defaulting non-investment-grade (NIG) issues, which reflect higher expected default probability for high yield ex-ante.[6] For IG and NIG issues that default without post-default prices, we use the corresponding IG and NIG default return averages as proxies for default-month returns. Using the in-sample composite default-month returns for defaulting bonds of similar credit quality, but without valid post-default pricing information, enables us to avoid the delisting bias shown in previous research on equity returns (Shumway, 1997).

3.2 Corporate Bond and Equity Characteristics

We build a comprehensive data library of 43 corporate bond characteristics that are either theoretically motivated or empirically identified by earlier studies on the cross-section of corporate bond returns. This broad set of bond return predictors can be largely classified into (i) bond-level characteristics such as issuance size, age, credit rating, time-to-maturity, and duration, (ii) proxies of corporate bond downside risk, (iii) proxies of bond-level illiquidity and liquidity risk, (iv) proxies of systematic risk such as default and term betas and volatility betas, (v) past bond return characteristics such as bond momentum, short-term reversal, and long-term reversal, and (vi) distributional characteristics including return volatility, skewness, and kurtosis. Appendix B provides a detailed description of these 43 bond characteristics as well as the studies that we follow closely to construct these measures. This list of corporate bond characteristics is not an exhaustive analysis of all possible predictors of corporate bond returns. Nonetheless, our list is designed to be representative of a broad set of corporate bond characteristics motivated in the literature for their explanatory power for bond returns. For equity characteristics, we rely on a large set of 94 stock-

[6] Consistent with Bali, Subrahmanyam, and Wen (2021a) who use a common dataset of bond returns after July 2002, the frequency of default events is rare in our sample.

level predictors used by Green, Hand, and Zhang (2017).[7] We restrain our equity characteristics sample to begin from July 2002 and end in December 2017 because we focus on the common sample period when our bond returns and characteristics become available in TRACE which starts in July 2002.

Our final sample includes 22,980 bonds issued by 1,841 unique firms, yielding a total of 146,085 firm-level bond-month return observations during the sample period from July 2002 to December 2017. Panel A of Table 1 reports the time-series average of the cross-sectional bond returns' distribution and bond characteristics. The numbers are presented at the firm-level using value-weighted average of firm-level bond returns and bond characteristic measures. The sample contains bonds with an average rating of 10.08 (i.e., BBB-), an average issue size of \$500 million, and an average time-to-maturity of 8.05 years. Among the full sample of bonds, about 75% are investment-grade and the remaining 25% are high-yield bonds. Panel B of Table 1 presents the correlation matrix for some of the firm-level bond characteristics and risk measures. As shown in Panel B, downside risk (i.e., proxied by the 5% Value-at-Risk) is positively associated with bond market beta (β^{Bond}), illiquidity, and rating, with respective correlations of 0.61, 0.19, and 0.25. The bond market beta, β^{Bond}, is also positively associated with rating and illiquidity, with respective correlations of 0.01 and 0.04. Bond maturity and duration are positively correlated with most risk measures, implying that bonds with longer maturity or duration (i.e., higher interest rate risk) have higher β^{Bond} and higher ILLIQ. Bond size is negatively correlated with ILLIQ, indicating that bonds with smaller size have higher ILLIQ.

4 Predicting Bond Returns without Hedge Ratios

4.1 Using only Bond Characteristics

We start our analysis with the baseline scenario of predicting bond returns without imposing cross-asset restrictions and using bond characteristics. Using the notation from equation (12) of Section 2.2, our goal in this subsection is to predict corporate bond returns as $E_t(RB_{it+1}) = f_1(XB_{it})$.

4.1.1 Out-of-Sample Predictive Power

Table 2 presents the monthly R^2_{OS} (in percentage) for the entire pooled sample of corporate bonds using all the 43 bond characteristics as covariates. The results are presented at the firm-level in

[7] Details on each of the 94 equity characteristics can be found in the Appendix in Green, Hand, and Zhang (2017) and Table A.6 in Gu, Kelly, and Xiu (2020). For missing equity (bond) characteristics, we follow Gu, Kelly, and Xiu (2020) and replace them with the cross-sectional median characteristic of each stock (bond) for each month.

Table 2 using value-weighted average of firm-level bond returns. Panel A of Table 2 reports R^2_{OS} for the entire sample of corporate bonds. The first column shows that the OLS model with all 43 bond characteristics produces an R^2_{OS} of -3.36%, indicating that the model fails to deliver significant out-of-sample forecasting power for the expected corporate bond returns. However, the other columns of Table 2 show that the machine learning models substantially improve the R^2_{OS}.[8] For example, by forming a few linear combinations of predictors via dimension reduction, columns (2) and (3) of Table 2 show that PCA and PLS improve the R^2_{OS} to 2.07% and 2.03%, respectively. By introducing the penalized methods into the loss function, columns (4) to (6) show that LASSO, Ridge, and ENet approach improve the R^2_{OS} to 1.85%, 1.89%, and 1.87%, respectively.

Unlike the linear models in column (1), regression trees are fully nonparametric and can reduce overfitting in individual bootstrap samples, and make the predictive performance more stable. Consistent with this prediction, column (7) of Table 2 shows a significant increase in R^2_{OS} to 2.19% using random forests (RF). In addition to nonparametric regressions, we investigate the performance of different neural network models including the feed forward neural networks (FFN) and the long short-term memory neural network (LSTM). As a typical neural network, feed forward neural networks (FFN) produces more flexible prediction approach by adding hidden layers between the inputs and output layer that aggregates hidden layers into the outcome prediction. The long short-term memory neural network (LSTM) captures long-term dependencies as a flexible hidden state space model for a large dimensional system. Columns (8) and (9) show that the FFN and LSTM models produce significant R^2_{OS} values of 2.37% and 2.28%, respectively. Finally, the last column of Table 2 shows that the forecast combination model (Combination) significantly improves the R^2_{OS} to 2.09%.[9]

To make pairwise comparisons of the estimation methods, we use the Diebold and Mariano (1995) test for differences in out-of-sample predictive accuracy between two models. Panel B of Table 2 reports the Diebold-Mariano test statistics for pairwise comparisons of a column model versus a row model. A positive statistic indicates that the column model outperforms the row model. The first row of Panel B shows a positive and statistically significant test statistic for all the machine learning models with Diebold-Mariano test statistics ranging from 2.89 to 3.85, compared to the unconstrained OLS model. Thus, all machine learning methods produce statistically significant improvements over the unconstrained OLS model. Comparisons between machine learning methods

[8]In Table 2, all of the R^2_{OS} statistics for the machine learning models are statistically significant with p-values less than 1%.

[9]Despite significant improvements in the forecasting performance, the R^2_{OS} of 2.09% based on the Combination model is slightly lower than those of RF (2.19%), FFN (2.37%), and LSTM (2.28%) models. This is plausible because the mean squared forecast error (MSFE) can be decomposed into forecast variance and the squared forecast bias (Rapach, Strauss, and Zhou, 2010) so that a model's forecasting performance depends on the tradeoff between the reduction in variance and bias. Combination model may significantly reduce the forecast variance but increases the bias of estimation, whereas the individual machine learning models such as RF and FFN may deliver better performance due to their ability to further reduce the forecasting biases which outweigh the costs of increasing variance.

themselves show that there is little difference in the performance of dimension reduction methods (PCA and PLS), penalized linear methods (LASSO, Ridge, ENet, and RF), and neural networks (FFN and LSTM), as the test statistics are not significant. Finally, the last column of Panel B shows that the forecast combination model (Combination) produces large and statistically significant improvements over most individual machine learning models.

4.1.2 Which Bond Characteristics Matter?

Next, we identify the corporate bond characteristics that are important determinants of the expected bond returns while simultaneously controlling for the many other predictors. We take the value-weighted average of bond-level characteristics to generate the firm-level bond characteristic measures. Following the ranking approach in Kelly, Pruitt, and Su (2019) and Gu, Kelly, and Xiu (2020), we discover influential covariates from setting all values of predictor j to zero, while holding the remaining model estimates fixed. The variable importance of the j^{th} input variable is measured by the reduction in panel prediction R^2_{OS}, which allows us to investigate the relative importance of individual bond characteristics for the performance of each machine learning model. To begin, for each of the nine machine learning methods, we calculate the reduction in R^2_{OS} from setting all values of a given predictor to zero within each training sample, and then average these into a single importance measure for each predictor. Figure 1 reports the resulting forecasting performance of the top 10 bond-level characteristics for each method, whereas Figure 2 reports overall rankings of characteristics for all models.[10]

Figures 1 and 2 demonstrate that all machine learning models are generally in close agreement regarding the most influential bond-level characteristics, which can be classified into four categories (i) bond characteristics related to interest rate risk such as duration (DUR) and time-to-maturity (MAT), (ii) risk measures such as downside risk proxied by Value-at-Risk (VaR) and expected shortfall (ES), total return volatility (VOL), and systematic risk related to bond market beta, default beta, term beta, and economic uncertainty beta (β^{Bond}, β^{DEF}, β^{TERM}, and β^{UNC}), (iii) bond-level illiquidity measures such as the average bid and ask price (AvgBidAsk), Amihud and Roll's measures of illiquidity, and (iv) past return characteristics related to bond momentum (MOM), short-term reversal (STR), and long-term reversal (LTR). Figure 1 shows that the risk measures play an important role in the dimension reduction methods (PCA and PLS), whereas bond-level characteristics related to interest rate risk are more prominent in the penalized methods (Lasso, Ridge, and Enet). Regression trees such as the random forest model rely more heavily on bond-level illiquidity measures such as the average bid and ask price and the Amihud measure. Neural networks such as FFN and LSTM draw predictive information mainly from bond return characteristics such as bond momentum and short-term reversal. Finally, the forecast combination

[10]The color gradient within each column in Figure 2 shows the model-specific ranking of characteristics, where the lightest (darkest) color indicates the least (most) important bond characteristics for each model.

model shows that bond momentum (MOM), return volatility (VOL), coskewness (COSKEW), and illiquidity (ILLIQ) are the top important covariates for the predictive performance.

In addition to comparing the covariate importance across all 43 firm-level bond characteristics, we further investigate their importance within each of the four characteristic groups. Panel A of Figure 3 shows that time-to-maturity is the most important characteristic for the expected bond returns within the Group I characteristics for all models, followed by duration. Panel B shows that within the Group II characteristics, coskewness and downside risk measures including VaR and ES, and systematic risk such as the macroeconomic uncertainty beta and default beta are the most important covariates. Panel C shows that the illiquidity measures such as the average bid and ask price play an important role across all machine learning models, whereas Panel D shows that higher return moments such as VOL as well as past return characteristics related to bond momentum are the top important covariates.

To find out which one of the four characteristic groups is the most important determinant of the expected bond returns, we present the relative strength of the four characteristic groups, respectively, in Figure 4, which shows a 10×4 bar chart representing the importance of each characteristic group for all methods. The columns of Figure 4 correspond to individual models, and color gradients within each column present a ranking from the most influential (dark blue) to the least influential (white) characteristic group. Figure 4 shows that the top two most important determinants are the characteristics related to bond-level illiquidity and liquidity risk (i.e., Group III) and the risk measures such as downside risk and systematic risk proxies (i.e., Group II).

4.1.3 Machine Learning based Long-Short Portfolios

To further investigate the economic significance of the machine learning models, we form portfolios based on the machine learning forecasts using the 43 bond characteristics. At the end of each month, we calculate the one-month-ahead out-of-sample firm-level bond return predictions for each of the ten methods (including the OLS). We then sort firm-level bond returns into deciles based on each model's forecasts of the one-month-ahead returns and then construct the value-weighted long-short portfolios of corporate bonds.[11] Table 3 reports the monthly performance results. "Low" is the decile portfolio with the lowest one-month-ahead expected return forecast (decile 1), "High" is the decile portfolio with the highest one-month-ahead expected return forecast (decile 10), and "High−Low" denotes the long-short portfolio that buys the highest expected return bonds in decile 10 and sells the lowest expected return bonds in decile 1. The returns are in percent per month and Newey-West t-statistics are reported in parentheses in the last column.

Table 3 presents the firm-level bond return results from long-short portfolios. Consistent with

[11] Following Bai, Bali, and Wen (2019), we use the bond's outstanding dollar values as weights. Since our statistical objective functions minimize equally weighted forecast errors, we also repeat the analysis using the equal-weighted portfolios and obtain qualitatively similar results.

our earlier findings using R^2_{OS} as the performance metric, Table 3 shows that all machine learning forecasts generate economically and statistically significant return spreads on the long-short bond portfolios, in the range of 0.33% and 0.79% per month, compared to the unconstrained OLS model which delivers the smallest return spread of 0.16%. The top three best hedge portfolios are generated by the RF, FFN, and LSTM, with the monthly return spread of 0.79% (t-statistic = 2.78), 0.75% (t-statistic = 2.61), and 0.79% (t-statistic = 3.33), respectively. The forecast combination model (Combination) also generates economically and statistically significant return spread of 0.67% (t-statistic = 3.41). Overall, Table 3 shows that the machine learning approaches significantly improve the forecasting performance for bond portfolios using firm-level bond characteristics as the covariates.

In unreported results, we also calculate the alphas and their t-statistics for the four-factor model of Bai, Bali, and Wen (2019) with the aggregate corporate bond market, the downside risk, the credit risk, and the liquidity risk factors of corporate bonds. Consistent with the strong explanatory power of these factors in explaining the cross-sectional variation in bond returns, we find that none of the alpha spreads is statistically significant. This is not surprising given that downside risk, credit risk, and liquidity risk as a whole are known to be pervasive and strong determinants of the expected bond returns.

4.2 Using Stock Characteristics

Equity and corporate bonds are contingent claims on firm fundamentals but also differ in several key features such as the payoff structure and the markedly different institutional and informational frictions across equities and bonds. Motivated by these observations, a few studies investigate whether a variety of stock characteristics predict corporate bond returns using cross-sectional Fama-MacBeth regressions (Chordia et al., 2017; Choi and Kim, 2018). These studies find mixed evidence on the role of stock characteristics for predicting future bond returns.[12] Compared to these studies which draw from the well of a limited number of predictors, we extend the list to a much larger set of stock characteristics and more importantly, we rely on machine learning methods to reduce redundant variation among predictors that address overfitting bias. Using the notation from equation (12) of Section 2.2, our goal in this subsection is to predict corporate bond returns as $E_t(RB_{it+1}) = f_1(XS_{it})$. In other words, while we do not impose the Merton (1974) model restrictions explicitly, we do allow for linkages between stock and bond returns in allowing stock characteristics (predictors of stock returns) to predict bond returns.

[12] For example, Chordia et al. (2017) find that many equity characteristics, such as accruals, standardized unexpected earnings, and idiosyncratic volatility, do not impact bond returns, whereas profitability and asset growth are negatively related to corporate bond returns. In contrast, Choi and Kim (2018) find that some variables (e.g., profitability and net issuance) fail to explain bond returns, and for others (e.g., investment and momentum) bond return premia are too large compared with their loadings, or hedge ratios, on equity returns of the same firms.

Table 4 presents the R^2_{OS} for the entire pooled sample of corporate bonds using all 94 stock characteristics from Green, Hand, and Zhang (2017) and Gu, Kelly, and Xiu (2020) as the covariates. The results in Table 4 are presented at the firm-level by constructing value-weighted average of firm-level bond returns, as well as the firm-level value-weighted average of bond characteristics, using amount outstanding as weights. Panel A of Table 4 shows that the OLS model with all 94 stock characteristics produces an R^2_{OS} of -3.09%, indicating that the model fails to deliver statistically significant out-of-sample forecasting power for the expected corporate bond returns. However, the other columns of Panel A show that the machine learning models substantially improve the R^2_{OS}. The penalized methods approach (LASSO, Ridge, and ENet) generate an R^2_{OS} of 1.61%, 1.57%, and 1.62%, respectively, all of which are similar to those delivered by the dimension reduction approach (PCA an PLS). Neural networks such as FFN and LSTM deliver significantly positive performance and improve the R^2_{OS} to 1.88% and 2.00%, respectively. Figure 5 plots the R^2_{OS} associated with stock characteristics and shows that the R^2_{OS} is in the range of 1.85% (Lasso) to 2.37% (FFN), which is in similar magnitude to those generated by using corporate bond characteristics, also presented in Figure 5.

In Panel B of Table 4, we form the long-short bond portfolios based on the machine learning forecasts using stock characteristics only (XS). Consistent with our earlier findings using the out-of-sample R-squared as the performance metric, Panel B shows that all machine learning forecasts generate economically and statistically significant return spreads on the long-short bond portfolios, in the range of 0.24% and 0.52% per month, compared to the unconstrained OLS model which delivers the smallest return spread of 0.02% (t-statistic = 0.12). Overall, Table 4 shows that the machine learning approaches significantly improve the return prediction performance for bond portfolios using the stock characteristics as the covariates.

4.3 Do Stock Characteristics Improve the Predictive Power of Bond Characteristics for Future Bond Returns?

The results so far suggest that all machine learning models produce significantly positive predictive power using either set of characteristics, and the predictive performance with using the bond characteristics is similar to that using the stock characteristics. In this section, we test whether the stock characteristics provide incremental power in predicting future bond returns relative to the bond characteristics. We start by predicting corporate bond returns as $E_t(RB_{it+1}) = f_1(XB_{it}, XS_{it})$.

Panel A of Table 5 reports R^2_{OS} from alternative estimation methods implemented with combining the 43 bond characteristics and 94 stock characteristics. Consistent with our previous findings, the traditional OLS model produces an R^2_{OS} of -5.38%, indicating that the model fails to deliver statistically significant out-of-sample forecasting power for the expected corporate bond returns. The other columns in Table 5, Panel A, show that the machine learning models using the combined 137 characteristics deliver significantly positive R^2_{OS} ranging from 1.60% (Ridge) to 2.11% (LSTM).

In Panel B, Table 5, we examine the improvement in the predictive power by comparing the machine learning bond portfolios formed based on the 137 characteristics, $f_1(XB, XS)$, to those formed using the 43 bond characteristics, $f_1(XB)$ from Section 4.1, or the 94 stock characteristics, $f_1(XS)$ from Section 4.2. Specifically, we calculate the difference in the High−Low long-short portfolio that takes a long position in the highest expected return bonds and a short position in the lowest expected return bonds based on different kinds of forecasts.

As shown in the last two rows of Panel B, Table 5, the economic significance of using both bond and stock characteristics is small compared to using bond characteristics alone. We find that most machine learning forecasts fail to deliver significantly positive return spread, indicating that there is no difference in the performance of the machine learning models when adding stock characteristics to the bond characteristics in forecasting future bond returns. In contrast, the last row of Panel B shows that most of the models deliver significantly positive return spreads, indicating the improvement in the models' performance when adding bond characteristics to the stock characteristics in predicting future bond returns. Overall, we conclude that although stock characteristics produce significant explanatory power for bond returns when used alone, their incremental predictive power relative to bond characteristics is economically insignificant, whereas bond characteristics play a major role and improve the performance of stock characteristics in predicting future bond returns.

4.4 Robustness Checks

4.4.1 Transaction Cost

Table OA1 of the Online Appendix provides robust checks of the main results in Table 3 and reports the monthly performance of value-weighted decile portfolios sorted on out-of-sample machine learning return forecasts using the 43 bond characteristics after accounting for transaction costs. Following Bao, Pan, and Wang (2011), we use the Roll (1984) measure of effective spreads calculated from autocovariances of bond returns and calculate transaction costs as the product of the portfolio turnover and the time-series mean of the cross-sectional average effective spread. Consistent with the findings in Table 3, Table OA1 shows that the machine learning approaches provide significantly positive long-short portfolio returns net of transactions costs.[13]

[13] A relatively low transaction cost is mainly driven by a low portfolio turnover, due to the persistence of predicted bond returns.

4.4.2 Time-varying Performance

We investigate the time-varying performance of the machine learning bond portfolio returns generated in Table 3. Table OA2 of the Online Appendix provides robustness checks and reports the conditional portfolio performance across different economic states based on the Chicago Fed National Activity Index (CFNAI).[14] The results in Table OA2 show that the machine learning bond portfolios exhibit significantly positive returns in both states of the economy, whereas the unconstrained OLS model delivers insignificant return spread of 0.14% (t-statistic = 1.38) and 0.11% (t-statistic = 1.32) in good and bad economic state, respectively.

4.4.3 Maturity-matched Bond Excess Returns

Throughout the paper we measure bond excess return as the difference between bond return and the risk-free rate proxied by the one-month Treasury bill rate. Table OA3 of the Online Appendix provides robust checks of the main results in Table 3 using maturity-matched Treasury returns to calculate bond excess returns. Consistent with the findings in Table 3, Table OA3 shows that the machine learning approaches provide significantly positive long-short portfolio returns after accounting for maturity-matched Treasury returns, with return spreads in the range of 0.31% and 0.73% per month.

4.4.4 Removing Financial Firms

We investigate whether our results are sensitive to the exclusion of financial firms. Following Fama and French (1992), we exclude financial firms with SIC codes between 6000 and 6999 because the high leverage that is normal for these firms probably does not have the same implication for non-financial firms, where high leverage is more likely to indicate financial distress. Consistent with our earlier findings, Table OA4 of the Online Appendix replicates the main findings in Table 2 (Panel A), Table 4 (Panel B), Table 5 (Panel C), Table 6 (Panel D), Table 7 (Panel E), Table 8 (Panel F), and Table 9 (Panel G) and shows similar results.

[14] The CFNAI is a monthly index designed to assess overall economic activity and related inflationary pressure. The CFNAI is a weighted average of 85 existing monthly indicators of national economic activity. It is constructed to have an average value of zero and a standard deviation of one. An index value below (above) zero corresponds to a good (bad) economic state.

5 Predicting Bond Returns with Regression-Based Hedge Ratios

We have so far shown that the marginal improvement of the forecasting power of stock characteristics relative to bond characteristics is economically small and statistically insignificant in predicting future bond returns. The results of the previous section thus seem to provide prima facie evidence of segmentation in the two markets. However, as noted in Section 2.2, the approach in the previous section is a reduced-form approach that does not explicitly link the functional forms of bond and stock expected return. In this section, we impose the dependence between expected bond and stock return via Merton (1974) model and investigate the incremental power of stock characteristics for future bond returns. The steps involved in estimating equation (13) are as follows:

1. Estimate hedge ratios via regressions. Following Choi and Kim (2018), our baseline estimate of the hedge ratio ($\hat{h}_{it}$) is based on the 36-month rolling window regression,

$$RB_{is} = \alpha_i + h_{it} RS_{is} + eB_{is}, \quad s = t-35, \ldots, t, \tag{19}$$

   where RB_{is} is the firm-level excess bond returns in month s and RS_{is} is the excess equity return of the same firm i in month s. The output is $\hat{h}_{it}$.

2. Calculate $RBmRS_{it+1} = RB_{it+1} - \hat{h}_{it} \times RS_{it+1}$, following equation (10).

3. Run separate machine learning models to predict the expected stock return $E_t(RS_{it+1})$ and $RBmRS_{it+1}$.

$$\begin{aligned} E_t(RBmRS_{it+1}) &= \psi_2(XB_{it}) \\ E_t(RS_{it+1}) &= \psi_1(XS_{it}). \end{aligned} \tag{20}$$

   Taking $E_t(RBmRS_{it+1}) = \psi_2(XB_{it})$ as an example, the machine learning model is given bond characteristics (XB_{it}) inputs to forecast the "dependent variable" ($RBmRS_{it+1}$). The output is a number $\psi_2(XB_{it})$ for each firm i and month t.

4. The prediction for expected bond return, a function of stock and bond characteristics and the hedge ratio, is then given by plugging in the estimated quantities in equation (13) to obtain:

$$E_t(RB_{it+1}) = f_2(XB_{it}, XS_{it}, \hat{h}_{it}) = \hat{h}_{it} \times \psi_1(XS_{it}) + \psi_2(XB_{it}). \tag{21}$$

We then compare the forecasted bond returns in equation (21) to the ones from the previous Section 4 without hedge ratios. We consider predictions using only bond characteristics, $f_1(XB)$

from Section 4.1, and using both bond and stock characteristics, $f_1(XB, XS)$ from Section 4.3, and evaluate whether or not $f_2(XB, XS, \hat{h})$ significantly outperforms $f_1(\cdot)$.[15]

Table 6 presents the forecasted bond returns based on equation (21). Consistent with our earlier findings using R^2_{OS} as the performance metric, Panel A shows that all machine learning forecasts generate economically and statistically significant R^2_{OS}, in the range of 1.93% (LASSO) to 4.95% (Combination). Panel B of Table 6 reports the Diebold-Mariano test statistics for comparisons of $f_2(XB, XS, \hat{h})$ versus $f_1(XB)$ and $f_1(XB, XS)$. We find a positive and statistically significant test statistic for six of the nine machine learning models with Diebold-Mariano test statistics ranging from 0.21 (PCA) to 2.86 (Combination), compared to the bond return forecasts using only bond characteristics, $f_1(XB)$. Finally, the last row of Panel B shows a positive and statistically significant test statistic for all machine learning models, indicating superior performance of $f_2(XB, XS, \hat{h})$ compared to bond return forecasts generated using the combined stock and bond characteristics, $f_1(XB, XS)$.

To further investigate the economic significance of our findings, we form the long-short bond portfolios based on the machine learning forecasts based on equation (21). Consistent with our earlier findings using the R^2_{OS} as the performance metric, Table 7 shows that $f_2(XB, XS, \hat{h})$ generates economically and statistically significant return spreads on the long-short bond portfolios, in the range of 0.55% and 0.92% per month, compared to the unconstrained OLS model which delivers the smallest return spread of 0.18% (t-statistic = 1.07). Finally, the last two rows of Table 7 examine the improvement in the predictive power by comparing the machine learning bond portfolios formed based on the restrictions to those without restrictions. Specifically, we calculate the average return (double) differences of the machine learning High−Low bond portfolios, (i) formed from sorting on forecasts $f_2(XB, XS, \hat{h})$ and those on $f_1(XB)$, and (ii) formed from sorting on forecasts $f_2(XB, XS, \hat{h})$ and those on $f_1(XB, XS)$. As shown in Table 7, the average return differences of the machine learning bond portfolios are all economically large and statistically significant, indicating that there is improvement in the performance of the machine learning models when we impose restrictions from the Merton (1974) model. Overall, we conclude that it is important to impose such restrictions when estimating bond expected returns, where equity characteristics provide significant improvement above and beyond bond characteristics for future bond returns.

[15] Step 3 above involves predicting stock returns using stock characteristics. Gu, Kelly, and Xiu (2020) find that machine learning offers an improved description of expected return relative to traditional methods in forecasting future stock returns. Consistent with their findings, Table OA5 of the Online Appendix shows that the machine learning methods provide strong forecasting power using the stock characteristics.

6 Predicting Bond Returns with Machine Learning-Based Hedge Ratios

The previous section uses firm-specific hedge ratios estimated via regressions. Since we estimate rolling window regressions, the hedge ratios are allowed to vary over time. An alternative approach is to use a full-scale structural model and use the estimated parameters to calculate hedge ratios. For example, Schaefer and Strebulaev (2008) is the first article to provide a comprehensive investigation of the magnitude and statistical significance of the hedge ratio. In this section, we follow the spirit of their approach by estimating hedge ratio using different machine learning approaches. The hedge ratio estimated in this section is time-varying and also a function of bond characteristics, that is, $h_{it} = \phi_3(XB_{it})$. The steps involved in estimating equation (14) are as follows:

1. Estimate the hedge ratio, $\hat{h}(XB_{it})$, based on the following functional form using a 36-month rolling window:
$$RB_{is} = h(XB_{is-1})RS_{is} + uB_{is}, \quad s = t-35, \ldots, t, \tag{22}$$
where $RB_{i,s}$ is the firm-level excess bond returns in month s, calculated as the value-weighted average excess returns of individual bonds issued by firm i, and $RS_{i,s}$ is the excess equity return of the same firm i in month s. The machine is given inputs including the bond characteristics (XB_{is-1}), realized bond returns (RB_{is}), and realized stock returns (RS_{is}). The machine outputs a "fitted value" $\hat{E}(RB_{is}|XB_{is-1}, RS_{is}) = \hat{h}(XB_{is-1}) \times RS_{is}$, which could be linear or non-linear, depending on the specific machine learning model used. Using the outputs of the machine we can calculate both the out-of-sample fitted value $\hat{h}(XB_{it}) \times RS_{it+1}$ and the hedge ratio $\hat{h}(XB_{it}) = \phi_3(XB_{it})$.[16]

2. Calculate $RBmRS_{it}$ following equation (10) as
$$RBmRS_{it+1} = RB_{it+1} - \hat{h}(XB_{it}) \times RS_{it+1}.$$

3. Run separate machine learning models to predict RS_{it+1} and $RBmRS_{it+1}$.
$$\begin{aligned} E_t(RBmRS_{it+1}) &= \phi_2(XB_{it}) \\ E_t(RS_{it+1}) &= \phi_1(XS_{it}). \end{aligned} \tag{23}$$

[16] As an illustrative example, consider prediction using neural network. Given XB_{is-1}, the machine generates K units of neurons for each layer l as $XB_l^{(k)} = g\left(\theta_l^{(k)} XB_{is-1}\right)$, where $g(\cdot)$ is the nonlinear activation function. Then, in the last layer, we multiply each $XB_L^{(k)}$ by RS_{is}. The output is the fitted value $\widehat{RB}_{is} = g\left(XB_L^{(k)} RS_{is}\right) = \hat{h}(XB_{is-1}) \times RS_{is}$. We can calculate the out-of-sample fitted value as $\hat{E}(RB_{it+1}|XB_{it}, RS_{it+1}) = \hat{h}(XB_{it}) \times RS_{it+1}$. When needed, the hedge ratio itself can be recovered by 'setting' the stock return to be one to obtain $\hat{h}(XB_{it}) = \hat{E}(RB_{it+1}|XB_{it}, 1)$.

Taking $E_t(RBmRS_{it+1}) = \phi_2(XB_{it})$ as an example, the machine learning model is given inputs including the bond characteristics (XB_{it}) to forecast the "dependent variable" ($RBmRS_{it+1}$). The output is a number $\phi_2(XB_{it})$ for each firm i and month t. Note that the prediction of the stock return in equation (23) is the same as that in equation (20), $\phi_1(XS_{it}) = \psi_1(XS_{it})$.

4. The prediction for expected bond return, a function of stock and bond characteristics and the hedge ratio, is then given by plugging in the estimated quantities in equation (14) to obtain:

$$E_t(RB_{it+1}) = f_3(XB_{it}, XS_{it}, \hat{h}(XB_{it})) = \phi_3(XB_{it}) \times \phi_1(XS_{it}) + \phi_2(XB_{it}). \tag{24}$$

We then compare the forecasted bond returns from equation (24), $f_3(XB, XS, \hat{h}(XB))$, to $f_2(XB, XS, \hat{h})$ from Section 5, and evaluate whether or not forecasts from machine learning based hedge ratios significantly outperform bond returns forecasted using the regression-based hedge ratios.

Table 8 presents the forecasted bond returns based on equation (24). Consistent with our earlier findings using R^2_{OS} as the performance metric, Panel A shows that all machine learning forecasts generate economically and statistically significant R^2_{OS}, in the range of 2.04% to 5.70%. Panel B of Table 8 compares the forecasted bond returns with Merton model restriction and machine learning estimated hedge ratio (i.e., $f_3(XB, XS, \hat{h}(XB))$ to the bond return forecasts from Section 4 obtained using bond characteristics, $f_1(XB)$, the combined stock and bond characteristics, $f_1(XB, XS)$, and bond return forecasts from Section 5 using $f_2(XB, XS, \hat{h})$. Panel B shows a positive and statistically significant test statistic for all the machine learning models with Diebold-Mariano test statistics, compared to the bond return forecasts using only bond characteristics, $f_1(XB)$, or the combined stock and bond characteristics, $f_1(XB, XS)$. Finally, the last row of Panel B shows a positive and statistically significant test statistic for all machine learning models, indicating superior performance of $f_3(XB, XS, \hat{h}(XB))$ compared to bond return forecasts generated using regression-based hedge ratios, $f_2(XB, XS, \hat{h})$.

Table 9 investigates the long-short portfolios of corporate bonds constructed with the machine learning forecasts based on $f_3(XB, XS, \hat{h}(XB))$. Consistent with our earlier findings using the out-of-sample R-squared as the performance metric, Table 9 shows that $f_3(XB, XS, \hat{h}(XB))$ generates economically and statistically significant return spreads on the long-short bond portfolios, in the range of 0.54% and 1.00% per month, compared to the unconstrained OLS model which delivers the smallest return spread of 0.16% (t-statistic = 0.53). Moreover, the average return spreads on the machine learning bond portfolios are all economically large and statistically significant, indicating that there is improvement in the performance of the machine learning models when we impose restrictions from the Merton (1974) model with machine learning estimated hedge ratio. Finally, the last row of Table 9 shows small and insignificant return spreads, indicating a relatively small improvement in economic significance between $f_3(XB, XS, \hat{h}(XB))$ estimated in equation (24) and $f_2(XB, XS, \hat{h})$ estimated from equation (21). Overall, we conclude that machine learning based hedge ratios provide more accurate predictions than the regression-based hedge ratios in terms of statistical significance. However, the economic significance of the predictions from both approaches turns out to be similar. One possible reason is that regression-based hedge ratios, being calculated over rolling windows, already account for time-variation in hedge ratios. We investigate these hedge ratios next.

To what extent the regression-based and machine-learning-based hedge ratios differ from each other? To answer this question, we choose the stochastic variance-based hedge ratio, equation (3) of Section 2.1, as the benchmark, and compare the mean squared error (MSEs) of the regression-based hedge ratio with the machine-learning-based ones. Specifically, let h_{it} and $\hat{h}_{it}$ be the benchmark and alternative hedge ratios, respectively. The MSE can be defined as

$$\mathrm{MSE} = \mathrm{E}(\hat{h}_{it} - h_{it})^2. \tag{25}$$

To calculate the benchmark hedge ratio, we estimate the asset variance following equation (8) of Schaefer and Strebulaev (2008), with which we calculate the long-term mean (θ) and volatility of volatility (γ). We assume the speed of mean reversion $\kappa = 4$ across all firms.

Table 10 reports the MSEs of different hedge ratios. We find that the MSE for the regression-based hedge ratio is 0.051, similar to those delivered by the machine learning-based hedge ratios, in the range of 0.053 (Ridge) and 0.057 (FFN). Both the regression-based and machine learning-based hedge ratio MSEs are much smaller than the unconstrained OLS model, which delivers the highest MSE of 0.097. The next two rows in this table report the MSE for the subsample based on the firm-level credit rating of individual bonds, and show smaller MSEs for non-investment-grade bonds than investment-grade bonds. The last two rows of the table show the smallest MSE for short-maturity bonds compared to the medium- and long-maturity bonds. Overall, the results are consistent with our earlier findings in Section 6 that the economic significance of using machine learning-based hedge ratio is similar to that using regression-based hedge ratio.

7 Conclusion

Using a variety of machine learning methods, we provide a comprehensive study of the cross-sectional pricing of corporate bonds using a large set of 94 stock characteristics and 43 bond characteristics. Because of the nonlinear payoffs of corporate bonds and the high correlation between many of the stock and bond characteristics, machine learning approaches are well suited for such challenging prediction problems by mitigating overfitting biases and uncovering complex patterns and hidden relationships.

Motivated by the Merton (1974) model that both equity and corporate bonds are contingent claims on firms, we explicitly link the functional forms of bond and stock expected returns by imposing economic structure when investigating bond expected returns. We find that the traditional linear regression models such as the OLS perform poorly, whereas the machine learning methods substantially improve the out-of-sample performance in predicting the cross-sectional differences in future bond returns. We show that using the reduced-form approach, the incremental improvement of stock characteristics relative to bond characteristics is economically and statistically small in forecasting future bond returns. However, after imposing the dependence between expected returns of bonds and stocks via the Merton (1974) model, we find economically and statistically large improvement in all machine learning forecasting models compared to the ones without any restrictions. Overall, our work highlights the importance of explicitly imposing the dependence between expected bond and stock returns when investigating expected bond returns.

Appendices

A Derivation of h_t and α_t

This section provides the analytical solutions for h_t and α_t in Section 2.1.

When the variance of the firm value is constant, according to Merton (1974), the equity price is equal to the European call price:

$$E_t = C_t(\sigma^2) = V_t N(d_1) - Be^{-r(T-t)}N(d_2), \tag{A.1}$$

where $d_1 = \frac{\ln(V_t/B)+(r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$, $d_2 = d_1 - \sigma\sqrt{T-t}$, and $N(\cdot)$ is the standard normal cumulative distribution function. When the variance is stochastic, Hull and White (1987) propose to approximate E_t at $\mathbb{E}[\bar{\sigma}^2_{t,T}]$ with Taylor expansion as

$$E_t \approx C_t(\mathbb{E}[\bar{\sigma}^2_{t,T}]) + \frac{1}{2}\left.\frac{\partial^2 C_t(\sigma^2)}{\partial\sigma^2}\right|_{\sigma^2=\mathbb{E}[\bar{\sigma}^2_{t,T}]} \cdot \mathrm{Var}(\bar{\sigma}^2_{t,T}), \tag{A.2}$$

where $\bar{\sigma}^2_{t,T} = \frac{1}{T-t}\int_t^T \sigma_s^2 ds$ is average variance of stochastic variance over time t to maturity T. Given equation (A.1), together with Cox, Ingersoll, and Ross (1985), we have

$$\mathbb{E}[\bar{\sigma}^2_{t,T}] = \frac{1}{T-t}\int_t^T \mathbb{E}[\sigma_s^2|\sigma_0^2]ds = \theta + \frac{e^{-\kappa t}-e^{-\kappa T}}{\kappa(T-t)}(\sigma_0^2-\theta). \tag{A.3}$$

$$\begin{aligned}\mathrm{Var}(\bar{\sigma}^2_{t,T}) &= \frac{1}{T-t}\int_t^T \mathrm{Var}(\sigma_s^2|\sigma_0^2)ds \\ &= \frac{\theta\gamma^2}{2\kappa} + \frac{\gamma^2(e^{-\kappa t}-e^{-\kappa T})}{\kappa(T-t)}(\sigma_0^2-\theta) + \frac{\gamma^2(e^{-2\kappa t}-e^{-2\kappa T})}{4\kappa^2(T-t)}(\theta-2\sigma_0^2).\end{aligned} \tag{A.4}$$

In the literature of asset pricing with stochastic volatility, κ is sufficiently positive. For example, Aït-Sahalia and Kimmel (2007) suggest $\kappa > 4$ for pricing equity index options, which implies a fairly fast speed of σ_t^2 converting to its long-run mean θ. Thus, equations (A.3) and (A.4) can be approximated as

$$\mathbb{E}[\bar{\sigma}^2_{t,T}] = \theta, \tag{A.5}$$

$$\mathrm{Var}(\bar{\sigma}^2_{t,T}) = \frac{\theta\gamma^2}{2\kappa}. \tag{A.6}$$

With some algebra, we can rewrite equation (A.2) as

$$E_t = V_t N(d_1) - Be^{-r(T-t)}N(d_2) - \frac{\sqrt{\theta}\gamma^2}{8\kappa}\cdot\eta_t, \tag{A.7}$$

where

$$\eta_t = V_t\phi(d_1)\sqrt{T-t}\left[d_1 \cdot \frac{(T-t)-\frac{2}{\theta}\ln(\frac{V_t}{B})}{4\sqrt{\theta(T-t)}} + \frac{1}{2\theta}\right], \tag{A.8}$$

and $\phi(\cdot)$ is the probability density function of standard normal distribution. Clearly, equation (A.7) reduces to equation (A.1) when $\gamma = 0$ and $\sigma_t^2 = \theta$, the case with constant variance.

Now we derive the hedge ratio as follows. According to Schaefer and Strebulaev (2008), the ratio is defined as:

$$h_t = \left[\left(\frac{\partial E_t}{\partial V_t} \right)^{-1} - 1 \right] \frac{E_t}{D_t}. \tag{A.9}$$

With constant variance, from equation (A.7) we get:

$$h_t = \left[\frac{1 - N(d_1)}{N(d_1)} \right] \frac{E_t}{D_t}. \tag{A.10}$$

In this case, the systemic risk of bond returns can be perfectly hedged by equity returns (because both bond and equity are driven by the dynamics of firm value alone):

$$\frac{dD_t}{D_t} - h_t \frac{dE_t}{E_t} = 0. \tag{A.11}$$

In the case of stochastic variance, the hedge ratio of equation (A.9) is

$$h_t = \frac{1 - N(d_1) + \gamma^2 \zeta_t}{N(d_1) - \gamma^2 \zeta_t} \cdot \frac{E_t}{D_t}, \tag{A.12}$$

where

$$\zeta_t = \frac{\phi(d_1)}{8\kappa} \left[\frac{(T-t) - \frac{2}{\theta} \ln(\frac{V_t}{B})}{4\sqrt{\theta(T-t)}} \left(d_1^2 + d_1 \sqrt{\theta(T-t)} - 1 \right) + \frac{\sqrt{\theta(T-t)} - 2d_1}{2\theta} \right]. \tag{A.13}$$

Because bond and equity are driven by both the dynamics of firm value and variance, the systemic risk of bond returns cannot be perfectly hedged by equity returns, and they have the following relation:

$$\frac{dD_t}{D_t} - h_t \frac{dE_t}{E_t} = \alpha_t dt, \tag{A.14}$$

where

$$\alpha_t = \frac{\gamma^2 \phi(d_1) V_t}{8\kappa D_t} \cdot \frac{\sqrt{\theta(T-t)} \left[\delta_1 \delta_2 (1 - d_1^2) - \delta_2 d_1 / 2\theta + a_1 + a_2 \right]}{N(d_1) - \frac{\gamma^2}{8\kappa} \phi(d_1) \left[\sqrt{\theta(T-t)} \left(\delta_1 d_1 + \frac{1}{2\theta} \right) - \left(\delta_1 (1 - d_1^2) + \frac{d_1}{\theta} \right) \right]} \tag{A.15}$$

in which

$$d_1 = \frac{\ln(V_t/B) + (r + \theta/2)(T-t)}{\sqrt{\theta(T-t)}}, \qquad \delta_1 = \frac{(T-t) - 2/\theta \cdot \ln(V_t/B)}{4\sqrt{\theta(T-t)}}, \qquad \delta_2 = \frac{\frac{\ln(V_t/B)}{T-t} - (r + \frac{\theta}{2})}{2\sqrt{\theta(T-t)}},$$

$$a_1 = \frac{\sigma_t^2 \left[\delta_1 (1 - d_1^2) - d_1/\theta \right] - d_1/4}{2\sqrt{\theta(T-t)}}, \qquad a_2 = \frac{\sigma_t^2 \left[\delta_1 d_1 (3 - d_1^2) + \frac{3 + d_1^2}{2\theta} \right] - \theta d_1 \delta_1 - 1/2}{2\theta(T-t)}.$$

B Corporate Bond Characteristics

This section describes a broad set of the 43 corporate bond characteristics, designed to be representative of (i) bond-level characteristics such as issuance size, credit rating, time-to-maturity, and duration, (ii) proxies of risk such as bond systematic risk, downside risk, and credit risk, (iii) proxies of bond-level illiquidity constructed using daily and intraday transaction data and liquidity risk, (iv) past bond return characteristics such as bond momentum, short-term and long-term reversals, and the distributional characteristics such as return volatility.

1. **Credit rating (*Rating*).** We collect bond-level rating information from Mergent FISD historical ratings. All ratings are assigned a number to facilitate the analysis, for example, 1 refers to a AAA rating, 2 refers to AA+, ..., and 21 refers to CCC. Investment-grade bonds have ratings from 1 (AAA) to 10 (BBB−). Non-investment-grade bonds have ratings above 10. A larger number indicates higher credit risk, or lower credit quality. We determine a bond's rating as the average of ratings provided by S&P and Moody's when both are available, or as the rating provided by one of the two rating agencies when only one rating is available.

2. **Time-to-maturity (*MAT*).** The number of years to maturity.

3. **Issuance size (*Size*).** The natural logarithm of bond amount outstanding.

4. **Age (*Age*).** Bond age since the first issuance, in the number of years.

5. **Duration (*DUR*).** A bond's price sensitivity to interest rate changes, measured in years.

6. **Downside risk proxied by the 5% VaR (*VaR5*).** Following Bai, Bali, and Wen (2019), we measure downside risk of corporate bonds using VaR, which determines how much the value of an asset could decline over a given period of time with a given probability as a result of changes in market rates or prices. Our proxy for downside risk, 5% Value-at-Risk ($VaR5$), is based on the lower tail of the empirical return distribution, that is, the second lowest monthly return observation over the past 36 months. We then multiply the original measure by -1 for convenience of interpretation.[17]

7. **Downside risk proxied by the 10% VaR (*VaR10*).** This measure is defined as the fourth lowest monthly return observation over the past 36 months. We then multiply the original measure by -1 for convenience of interpretation.

8. **Downside risk proxied by the 5% Expected Shortfall (*ES5*).** An alternative measure of downside risk, "expected shortfall," is defined as the conditional expectation of loss given that the loss is beyond the VaR level. In our empirical analyses, we use the 5% expected shortfall (ES5) defined as the average of the two lowest monthly return observations over the past 36 months (beyond the 5% VaR threshold).

9. **Downside risk proxied by the 10% Expected Shortfall (*ES10*).** An alternative measure of downside risk, "expected shortfall," is defined as the conditional expectation of loss given that the loss is beyond the VaR level. In our empirical analyses, we use the

[17] Note that the original maximum likely loss values are negative since they are obtained from the left tail of the return distribution. After multiplying the original VaR measure by -1, a positive regression coefficient and positive return/alpha spreads in portfolios are interpreted as the higher downside risk being related to the higher cross-sectional bond returns.

10% expected shortfall ($ES10$) defined as the average of the four lowest monthly return observations over the past 36 months (beyond the 10% VaR threshold).

10. **Illiquidity (*ILLIQ*).** A bond-level illiquidity measure. We follow Bao, Pan, and Wang (2011) to construct the measure, which aims to extract the transitory component from bond price. Specifically, let $\Delta p_{itd} = p_{itd} - p_{itd-1}$ be the log price change for bond i on day d of month t. Then, $ILLIQ$ is defined as

$$ILLIQ = -Cov_t(\Delta p_{itd}, \Delta p_{itd+1}).$$

11. **Roll's daily measure of illiquidity (*Roll*).** As an alternative measure of bond-level illiquidity using daily bond returns, the Roll (1984) measure is defined as,

$$Roll = \begin{cases} 2\sqrt{-\mathrm{cov}(r_d, r_{d-1})} & \text{if } \mathrm{cov}(r_d, r_{d-1}) < 0, \\ 0 & \text{otherwise,} \end{cases}$$

where r_d is the corporate bond return on day d. Given the fact that corporate bonds do not trade frequently, this measure crucially depends on two conditions. First, a bond is traded for two days in a row so that we can calculate its daily return. Second, a bond has at least a number of daily returns calculated each month so that we can calculate its covariance. We set the threshold equal to five. A bond's monthly *Roll* measure will be missing if that bond does not have five daily returns calculated that month.

12. **Roll's intraday measure of illiquidity (*TC_Roll*).** Following Dick-Nielsen, Feldhütter, and Lando (2012), we employ an intraday version of the Roll (1984) estimator for effective spreads,

$$TC\_Roll = \begin{cases} 2\sqrt{-\mathrm{cov}(r_i, r_{i-1})} & \text{if } \mathrm{cov}(r_i, r_{i-1}) < 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $r_i = \frac{P_i - P_{i-1}}{P_{i-1}}$ is the return of the ith trade.

13. **High-low spread estimator(*P_HighLow*).** Following Corwin and Schultz (2012), we use the ratio between the daily high and low prices on consecutive days to approximate bid-ask spreads. With such motivation, their effective spread proxy is defined as

$$\begin{aligned} P\_HighLow &= \frac{2(e^{\alpha} - 1)}{1 + e^{\alpha}}, \\ \alpha &= \frac{\sqrt{2\beta} - \sqrt{\beta}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}}, \\ \beta &= \sum_{j=0}^{1} \left(\ln \left(\frac{H_{t+j}}{L_{t+j}} \right) \right)^2, \\ \gamma &= \left(\ln \left(\frac{H_{t,t+1}}{L_{t,t+1}} \right) \right)^2. \end{aligned}$$

$H_t(L_t)$ is the highest (lowest) transaction price at day t, and $H_{t,t+1}(L_{t,t+1})$ is the highest (lowest) price on two consecutive days t and $t+1$. Again, we take the mean of the daily values in a month to get a monthly spread proxy for each bond.

14. **Illiquidity measure based on zero returns (*P_Zeros*).** Following Lesmond, Ogden,

and Trzcinka (1999), we use the proportion of zero return days as a measure of liquidity. Lesmond, Ogden, and Trzcinka (1999) argue that zero volume days (hence zero return days) are more likely to reflect lower liquidity. We compute their measure on a monthly basis with T as the number of trading days in a month,

$$P\_Zeros = \frac{\#\text{ of zero return days}}{T},$$

The number of zero return days comprises two parts, the sequential days with no price change hence zero returns, and the days with zero trading volume.

15. **Modified illiquidity measure based on zero returns (*P_FHT*)**. Fong, Holden, and Trzcinka (2017) propose a new bid-ask spread proxy based on the zeros measure in Lesmond, Ogden, and Trzcinka (1999). In their framework, symmetric transaction costs of $S/2$ leads to observed returns of

$$R = \begin{cases} R^* + \frac{S}{2} & \text{if } R^* < -\frac{S}{2}, \\ 0 & \text{if } -\frac{S}{2} < R^* < \frac{S}{2}, \\ R^* - \frac{S}{2} & \text{if } \frac{S}{2} < R^*, \end{cases}$$

where R^* is the unobserved true value return, which they assume to be normally distributed with mean zero and variance σ^2. Hence, they equate the theoretical probability of a zero return with its empirical frequency, measured via P_Zeros. Solving for the spread S, they get

$$P\_FHT = S = 2 \cdot \sigma \cdot \Phi^{-1}\left(\frac{1 + \text{P\_Zeros}}{2}\right)$$

where Φ^{-1} is the inverse of the cumulative standard normal distribution. We compute a bond's σ for each month and then calculate P_FHT.

16. **Amihud measure of illiquidity (*Amihud*)**. Following Amihud (2002), the measure is motivated to capture the price impact and is defined as,

$$Amihud = \frac{1}{N} \sum_{d=1}^{N} \frac{|r_d|}{Q_d},$$

where N is the number of positive-volume days in a given month, r_d the daily return, and Q_d the trading volume on day d, respectively.

17. **An extended Roll's measure (*PI_Roll*)**. Goyenko, Holden, and Trzcinka (2009) derive an extended transaction cost proxy measure, which for every transaction cost proxy tcp and average daily dollar volume $\overline{Q}$ in the period under observation is defined as

$$PI\_Roll = \frac{Roll}{\overline{Q}}.$$

18. **An extended FHT measure based on zero returns (*PI_FHT*)**.

$$PI\_FHT = \frac{P\_FHT}{\overline{Q}}.$$

where P_FHT is the modified illiquidity measure based on zero returns (Fong, Holden, and

Trzcinka, 2017) and $\overline{Q}$ is the average daily dollar volume in the period under observation.

19. **An extended High-low spread estimator (*PI_HighLow*).**

$$PI\_HighLow = \frac{P\_HighLow}{\overline{Q}}.$$

where $PI\_HighLow$ is the high-low spread estimator following Corwin and Schultz (2012) and $\overline{Q}$ is the average daily dollar volume in the period under observation.

20. **Std.dev of the Amihud measure (*Std_Amihud*).** The standard deviation of the daily Amihud measure within a month.

21. **Lambda (*PI_Lambda*).** Hasbrouck (2009) proposes Lambda as a high-frequency price impact measure for equities. PI_Lambda (λ) is estimated in the regression,

$$r_\tau = \lambda \cdot \text{sign}(Q_\tau) \cdot \sqrt{|Q_{\tau|}} + \epsilon_\tau,$$

where r_τ is the stock's return and Q_τ is the signed traded dollar volume within the five minute period τ. Following Hasbrouck (2009) and Schestag, Schuster, and Uhrig-Homburg (2016), we take into account the effects of transaction costs on small trades versus large trades (Edwards, Harris, and Piwowar, 2007) and run the adjusted regression,

$$r_i = \alpha \cdot D_i + \lambda \cdot D_i \cdot \sqrt{Q_i} + \epsilon_i,$$

where λ is estimated in the equation above excluding all overnight returns and D_i is an indicator variable of trades defined as the following,

$$D_i = \begin{cases} 1 & \text{if trade } i \text{ is a buy,} \\ 0 & \text{if trade } i \text{ is an interdealer trade,} \\ -1 & \text{if trade } i \text{ is a sell.} \end{cases}$$

22. **Difference of average bid and ask prices (*AvgBidAsk*).** Following Hong and Warga (2000) and Chakravarty and Sarkar (2003), we use the difference between the average customer buy and the average customer sell price on each day to quantify transaction costs:

$$AvgBidAsk = \frac{\overline{P_t^{Buy}} - \overline{P_t^{Sell}}}{0.5 \cdot (\overline{P_t^{Buy}} + \overline{P_t^{Sell}})}$$

where $\overline{P_t^{Buy/Sell}}$ is the average price of all customer buy/sell trades on day t. We calculate $AvgBidAsk$ for each day on which there is at least one buy and one sell trade and use the monthly mean as a monthly transaction cost measure.

23. **Interquartile range (*TC_IQR*).** Han and Zhou (2007) and Pu (2009) use the interquartile range of trade prices as a bid-ask spread estimator. They divide the difference between the 75th percentile P_t^{75th} and the 25th percentile P_t^{25th} of intraday trade prices on day t by the average trade price $\overline{P_t}$ of that day:

$$TC\_IQR = \frac{P_t^{75th} - P_t^{25th}}{\overline{P_t}},$$

We calculate TC_IQR for each day that has at least three observations and define the monthly measure as the mean of the daily measures.

24. **Round-trip transaction costs (*RoundTrip*)**. Following Feldhütter (2012), we aggregate all trades per bond with the same volumes that occur within a 15-minute time window to a round-trip transaction. We then compute the estimator for round-trip transaction costs as the doubled difference between the lowest and highest trade price for each round-trip transaction. To obtain a relative spread proxy, we divide the round-trip transaction cost estimator by the mean of the maximum and the minimum price. A bond's monthly round-trip measure is then obtained by averaging over all round-trip trades in a month.

25. **Pastor and Stambaugh's liquidity measure (*GammaPS*, γ_{PS})**. Pástor and Stambaugh (2003) develop a measure for price impact based on price reversals for the equity market. It is given by the estimator for γ in the following regression:

$$r_{t+1}^e = \theta + \psi \cdot r_t + \gamma \cdot sign(r_t^e) \cdot Q_t + \epsilon_t,$$

where r_t^e is the security's excess return over a market index return, r_t is the security's return and Q_t is the trading volume at day t. For corporate bond market index, we use Merrill Lynch aggregate corproate bond index. γ should be negative and a larger price impact leads to a larger absolute value. As liquidity measures generally assign larger (positive) values to more illiquid bonds, we define $\gamma_{PS} = -\gamma$ expect it to be positively correlated with the other liquidity measures.

26. **Bond market beta (β^{Bond})**. We estimate the bond market beta, β^{Bond}, for each bond from the time-series regressions of individual bond excess returns on the bond market excess returns (MKT^{Bond}) using a 36-month rolling window. We compute the bond market excess return (MKT^{Bond}) as the value-weighted average returns of all corporate bonds in our sample minus the one-month Treasury-bill rate.[18]

27. **Default beta (β^{DEF})**. We estimate the default beta for each bond from the time-series regressions of individual bond excess returns on the bond market excess returns (MKT^{Bond}) and the default factor using a 36-month rolling window. Following Fama and French (1993), the default factor (DEF) is defined as the difference between the return on a market portfolio of long-term corporate bonds (the composite portfolio on the corporate bond module of Ibbotson Associates) and the long-term government bond return.

28. **Term beta (β^{TERM})**. We estimate the default beta for each bond from the time-series regressions of individual bond excess returns on the bond market excess returns (MKT^{Bond}) and the term factor using a 36-month rolling window. Following Fama and French (1993), the term factor (TERM) is defined as the difference between the monthly long-term government bond return (from Ibbotson Associates) and the one-month Treasury bill rate.

29. **Illiquidity beta (β^{LWW})**. Following Lin, Wang, and Wu (2011), it is estimated as the exposure to the bond illiquidity factor, which is defined as the average return difference between the high liquidity beta portfolio (decile 10) and the low liquidity beta portfolio (decile 1).

[18] We also consider alternative bond market proxies such as the Barclays Aggregate Bond Index and Merrill Lynch Bond Index. The results from these alternative bond market factors turn out to be similar to those reported in our tables.

30. **Downside risk beta (β^{DRF}).** Following Bai, Bali, and Wen (2019), for each bond and each month in our sample, we estimate the factor beta from the monthly rolling regressions of excess bond returns on the downside risk factor (DRF) over a 36-month fixed window after controlling for the bond market factor (MKT^{Bond}).

31. **Credit risk beta (β^{CRF}).** Similar to the construction of downside risk beta, for each bond and each month in our sample, we estimate the factor beta from the monthly rolling regressions of excess bond returns on the credit risk factor (CRF) over a 36-month fixed window after controlling for the bond market factor (MKT^{Bond}).

32. **Illiquidity risk beta (β^{LRF}).** Similar to the construction of downside risk and credit risk beta, for each bond and each month in our sample, we estimate the factor beta from the monthly rolling regressions of excess bond returns on the liquidity risk factor (LRF) over a 36-month fixed window after controlling for the bond market factor (MKT^{Bond}).

33. **Volatility beta (β^{VIX}).** Following Chung, Wang, and Wu (2019), we estimate the following bond-level regression

    $$R_{i,t} = \alpha_i + \beta_{1,i}MKT_t + \beta_{2,i}SMB_t + \beta_{3,i}HML_t + \beta_{4,i}DEF_t + \beta_{5,i}TERM_t + \beta_{6,i}\Delta VIX_t + \epsilon_{i,t},$$

    where $R_{i,t}$ is the excess return of bond i in month t, and MKT_t, SMB_t, HML_t, DEF_t, $TERM_t$, and ΔVIX_t denote the aggregate corporate bond market, the size factor, the book-to-market factor, the default factor, the term factor, and the market volatility risk factor, respectively.

34. **Macroeconomic Uncertainty Beta (β^{UNC}).** Following Bali, Subrahmanyam, and Wen (2021b), for each bond-month in our sample, we estimate the uncertainty beta from monthly rolling regressions of excess bond returns on the change in the economic uncertainty index (ΔUNC) and the excess bond market returns (MKT), using the past 24 to 36 months of data (as available):

    $$R_{i,t} = \alpha_{i,t} + \beta_{i,t}^{UNC} \cdot \Delta UNC_t + \beta_{i,t}^{MKT} \cdot MKT_t + \epsilon_{i,t},$$

    where $R_{i,t}$ is the excess return of bond i in month t, ΔUNC_t is the change in the economic uncertainty index in month t based on Jurado, Ludvigson, and Ng (2015), MKT_t is the aggregate corporate bond market, $\beta_{i,t}^{UNC}$ is the uncertainty beta of bond i in month t.

35. **Short-term reversal (*REV*).** The bond return in previous month.

36. **Six-month momentum (*MOM6*).** Following Jostova et al. (2013), it is defined as the cumulative bond returns over months from $t-7$ to $t-2$ (formation period), skipping the short-term reversal month.

37. **Twelve-month momentum (*MOM12*).** It is defined as the cumulative bond returns over months from $t-12$ to $t-2$ (formation period), skipping the short-term reversal month.

38. **Long-term reversal (*LTR*).** Following Bali, Subrahmanyam, and Wen (2021a), it is defined as the past 36-month cumulative returns from $t-48$ to $t-13$, skipping the 12-month momentum and short-term reversal month.

39. **Volatility (*VOL*).** Following Bai, Bali, and Wen (2016), it is estimated using a 36-month

rolling window for each bond in our sample

$$VOL_{i,t} = \frac{1}{n-1}\sum_{t=1}^{n}(R_{i,t} - \overline{R}_i)^2.$$

40. **Skewness (*SKEW*).** Similar to the construction of volatility, skewness is estimated using a 36-month rolling window for each bond in our sample

$$SKEW_{i,t} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{R_{i,t} - \overline{R}_i}{\sigma_{i,t}}\right)^3.$$

41. **Kurtosis (*KURT*).** Similar to the construction of volatility and skewness, kurtosis is estimated using a 36-month rolling window for each bond in our sample

$$KURT_{i,t} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{R_{i,t} - \overline{R}_i}{\sigma_{i,t}}\right)^4 - 3.$$

42. **Co-skewness (*COSKEW*).** Harvey and Siddique (2000), Mitton and Vorkink (2007), and Boyer, Mitton, and Vorkink (2010) provide empirical support for the three-moment asset pricing models that stocks with high co-skewness, high idiosyncratic skewness, and high expected skewness have low subsequent returns. Following the aforementioned studies, we decompose total skewness into two components; systematic skewness and idiosyncratic skewness, which are estimated based on the following time-series regression for each bond using a 36-month rolling window:

$$R_{i,t} = \alpha_i + \beta_i \cdot R_{m,t} + \gamma_i \cdot R_{m,t}^2 + \varepsilon_{i,t}.$$

where $R_{i,t}$ is the excess return on bond i, $R_{m,t}$ is the excess return on the bond market portfolio, γ_i is the systematic skewness (co-skewness) of bond i.

43. **Idiosyncratic skewness (*ISKEW*).** The idiosyncratic skewness (*ISKEW*) of bond i is defined as the skewness of the residuals ($\varepsilon_{i,t}$) in co-skewness regression equation.

C Four Groups of Corporate Bond Characteristics

We classify all 43 corporate bond characteristics into the following four broad categories,

- Group I: Bond characteristics related to interest risk or maturity: *Rating*, *MAT*, *Size*, *Age*, *DUR*.
- Group II: Bond characteristics related to risk measures such as downside risk or systematic risk: *VaR5*, *VaR10*, *ES5*, *ES10*, β^{Bond}, β^{DEF}, β^{TERM}, β^{UNC}, β^{VIX}, β^{LWW}, β^{DRF}, β^{CRF}, β^{LRF}, *COSKEW*.
- Group III: Bond-level illiquidity and illiquidity risk: *ILLIQ*, *Roll*, *TCRoll*, *P_HighLow*, *P_Zeros*, *P_FHT*, *Amihud*, *PI_Roll*, *PI_FHT*, *PI_HighLow*, *Std_Amihud*, *PI_Lambda*, *AvgBidAsk*, *TC_IQR*, *Roundtrip*, γ_{PS}.
- Group IV: Past return characteristics: *REV*, *MOM6*, *MOM12*, *LTR*, *VOL*, *SKEW*, *KURT*, *ISKEW*.

References

Aït-Sahalia, Yacine, and Robert Kimmel, 2007, Maximum likelihood estimation of stochastic volatility models, *Journal of Financial Economics* 83, 413–452.

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Bai, Jennie, Turan G. Bali, and Quan Wen, 2016, Do the distributional characteristics of corporate bonds predict their future returns?, *Working Paper,* SSRN E-Library.

Bai, Jennie, Turan G. Bali, and Quan Wen, 2019, Common risk factors in the cross-section of corporate bond returns, *Journal of Financial Economics* 131, 619–642.

Bali, Turan G., Avanidhar Subrahmanyam, and Quan Wen, 2021a, Long-term reversals in the corporate bond market, *Journal of Financial Economics* 139, 656–677.

Bali, Turan G., Avanidhar Subrahmanyam, and Quan Wen, 2021b, The macroeconomic uncertainty premium in the corporate bond market, *Journal of Financial and Quantitative Analysis,* 56, 1653–1678.

Bao, Jack, Jun Pan, and Jiang Wang, 2011, The illiquidity of corporate bonds, *Journal of Finance* 66, 911–946.

Bessembinder, Hendrik, Kathleen M. Kahle, William F. Maxwell, and Danielle Xu, 2009, Measuring abnormal bond performance, *Review of Financial Studies* 22, 4219–4258.

Bessembinder, Hendrik, William F. Maxwell, and Kumar Venkataraman, 2006, Market transparency, liquidity externalities, and institutional trading costs in corporate bonds, *Journal of Financial Economics* 82, 251–288.

Black, Fischer, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.

Boyer, Brian, Todd Mitton, and Keith Vorkink, 2010, Expected idiosyncratic skewness, *Review of Financial Studies* 23, 169–202.

Breiman, Leo, 2001, Random forests, *Machine Learning* 45, 5–32.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone, 1984, Classification and regression trees Belmont, Calif.: Wadsworth.

Chakravarty, Sugato, and Asani Sarkar, 2003, Trading costs in three U.S. bond markets, *Journal of Fixed Income* 13, 39–48.

Chen, Luyang, Markus Pelger, and Jason Zhu, 2019, Deep learning in asset pricing. Working paper.

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

Chordia, Tarun, Amit Goyal, Yoshio Nozawa, Avanidhar Subrahmanyam, and Qing Tong, 2017, Are capital market anomalies common to equity and corporate bond markets?, *Journal of Financial and Quantitative Analysis* 52, 1301–1342.

Chung, Kee H., Junbo Wang, and Chunchi Wu, 2019, Volatility and the cross-section of corporate bond returns, *Journal of Financial Economics,* 133, 397–417.

Cici, Gjergji, Scott Gibson, and Rabih Moussawi, 2017, Explaining and benchmarking corporate bond returns, *Working Paper,* SSRN elibrary.

Clark, Todd E., and Kenneth D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291–311.

Cochrane, John H., 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

Corwin, Shane A., and Paul Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, *Journal of Finance* 67, 719–760.

Cox, John C., Jonathan E Ingersoll, and Stephen A. Ross, 1985, A theory of the term instructure of interest rates, *Econometrica* 53, 385–407.

Dick-Nielsen, Jens, Peter Feldhütter, and David Lando, 2012, Corporate bond liquidity before and after the onset of the subprime crisis, *Journal of Financial Economics* 103, 471–492.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 134–144.

Diebold, Francis X., and Minchul Shin, 2019, Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives, *International Journal of Forecasting* 35, 1679–1691.

Du, Du, Redouane Elkamhi, and Jan Ericsson, 2019, Time-varying asset volatility and the credit spread puzzle, *Journal of Finance* 74, 1841–1885.

Edwards, Amy K., Lawrence E. Harris, and Michael S. Piwowar, 2007, Corporate bond market transaction costs and transparency, *Journal of Finance* 62, 1421–1451.

Fama, Eugene F., and Kenneth R. French, 1992, Cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Feldhütter, Peter, 2012, The same bond at different prices: Identifying search frictions and selling pressure, *Review of Financial Studies* 25, 1155–1206.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the factor zoo: A test of new factors, *Journal of Finance* 75, 1327–1370.

Fong, Kingsley, Craig W. Holden, and Charles A. Trzcinka, 2017, What are the best liquidity proxies for global research?, *Review of Finance* 21, 1355–1401.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *Review of Financial Studies* 33, 2326–2377.

Gebhardt, William R., Soeren Hvidkjaer, and Bhaskaran Swaminathan, 2005, The cross section of expected corporate bond returns: betas or characteristics?, *Journal of Financial Economics* 75, 85–114.

Giglio, Stefano, Yuan Liao, and Dacheng Xiu, 2021, Thousands of alpha tests, *Review of Financial Studies* 34, 3456–3496.

Goyenko, Ruslan, Craig Holden, and Charles Trzcinka, 2009, Do liquidity measures measure liquidity?, *Journal of Financial Economics* 92, 153–181.

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang, 2017, The characteristics that provide independent information about average U.S. monthly stock returns, *Review of Financial Studies* 30, 4389–4436.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *Review of Financial Studies* 33, 2223–2273.

Han, Song, and Hao Zhou, 2007, Nondefault bond spread and market trading liquidity, *Working Paper* Federal Reserve Board.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263–1295.

Hasbrouck, Joel, 2009, Trading costs and returns for U.S. equities: Estimating effective costs from daily data, *Journal of Finance* 65, 1445–1477.

Hochreiter, Sepp, and Jürgen Schmidhuber, 1997, Long short-term memory, *Neural Computation* 9, 1735–1780.

Hong, Gwangheon, and Arthur Warga, 2000, An empirical study of bond market transactions, *Financial Analyst Journal* 56, 32–46.

Hong, Harrison, and David Sraer, 2013, Quiet bubbles, *Journal of Financial Economics* 110, 596–606.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

Hull, John, and Alan White, 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* 42, 281–300.

Jostova, Gergana, Stanislava Nikolova, Alexander Philipov, and Christof W. Stahel, 2013, Momentum in corporate bond returns, *Review of Financial Studies* 26, 1649–1693.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng, 2015, Measuring uncertainty, *American Economic Review* 105, 1177–1216.

Kelly, Bryan T., Diogo Palhares, and Seth Pruitt, 2022, Modeling corporate bond returns, *Journal of Finance,* forthcoming.

Kelly, Bryan T., Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross section, *Journal of Financial Economics* 135, 271–292.

Kwan, Simon H., 1996, Firm-specific information and the correlation between individual stocks and bonds, *Journal of Financial Economics* 40, 63–80.

Lesmond, David A., Joseph P. Ogden, and Charles A. Trzcinka, 1999, A new estimate of transaction costs, *Review of Financial Studies* 12, 1113–1141.

Lettau, Martin, and Markus Pelger, 2020, Factors that fit the time series and cross-section of stock returns, *Review of Financial Studies* 33, 2274–2325.

Lin, Hai, Junbo Wang, and Chunchi Wu, 2011, Liquidity risk and the cross-section of expected corporate bond returns, *Journal of Financial Economics* 99, 628–650.

Linnainmaa, Juhani T., and Michael R. Roberts, 2018, The history of the cross-section of stock returns, *Review of Financial Studies* 31, 2606–2649.

Lo, Andrew W., 1991, Long-term memory in stock market prices, *Econometrica* 59, 1279–1313.

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic publication destroy stock return predictability?, *Journal of Finance* 71, 5–32.

Merton, Robert C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449–470.

Mitton, Todd, and Keith Vorkink, 2007, Equilibrium underdiversification and the preference for skewness, *Review of Financial Studies* 20, 1255–1288.

Nagel, Stefan, 2021, Machine learning in asset pricing, Princeton University Press.

Pástor, Ľuboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Pu, Xiaoling, 2009, Liquidity commonality across the bond and cds markets, *Journal of Fixed Income* 19, 26–39.

Rapach, David E., Jack K. Strauss, and Guofu Zhou, 2010, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 23, 821–862.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 1127–1139.

Schaefer, Stephen M., and Ilya Strebulaev, 2008, Structural models of credit risk are useful: Evidence from hedge ratios on corporate bonds, *Journal of Financial Economics* 90, 1–19.

Schestag, Raphael, Philipp Schuster, and Marliese Uhrig-Homburg, 2016, Measuring liquidity in bond markets, *Review of Financial Studies* 29, 1170–1219.

Shumway, Tyler, 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327–340.

Table 1: Descriptive statistics

Panel A reports the total number of observations, the cross-sectional mean, median, standard deviation and monthly return percentiles of corporate bonds, and bond characteristics including credit rating, time-to-maturity (Maturity, year), amount outstanding (Size, $ million), duration, downside risk (5% Value-at-Risk, VaR), illiquidity (ILLIQ), and the CAPM beta based on the corporate bond market index, β^{Bond}. The numbers are presented at the firm-level using value-weighted average of firm-level bond returns and bond characteristic measures. Ratings are in conventional numerical scores, where 1 refers to an AAA rating and 21 refers to a C rating. Higher numerical score means higher credit risk. Numerical ratings of 10 or below (BBB- or better) are considered investment grade, and ratings of 11 or higher (BB+ or worse) are labeled high yield. Downside risk is the 5% Value-at-Risk (VaR) of corporate bond return, defined as the second lowest monthly return observation over the past 36 months. The original VaR measure is multiplied by -1 so that a higher VaR indicates higher downside risk. Bond illiquidity is computed as the autocovariance of the daily price changes within each month, multiplied by -1. β^{Bond} is the corporate bond exposure to the excess corporate bond market return, constructed using the Merrill Lynch U.S. Aggregate Bond Index. The betas are estimated for each bond from the time-series regressions of bond excess returns on the excess bond market return using a 36-month rolling window estimation. Panel B reports the time-series average of the cross-sectional correlations. The sample period is from July 2002 to December 2017.

Panel A: Cross-sectional statistics over the sample period of July 2002 – December 2017

					Percentiles			
	N	Mean	Median	SD	5th	25th	75th	95th
Bond return (%)	146,085	0.59	0.52	3.18	−3.04	−0.33	1.47	4.33
Rating	146,085	10.08	9.50	3.94	4.33	7.33	13.16	16.50
Time to maturity (maturity, year)	146,085	8.05	6.75	5.24	2.46	4.86	9.60	17.98
Amount out (size, $million)	146,085	500.17	390.00	413.65	150.00	250.00	594.64	1271.74
Duration (DUR)	146,085	4.63	4.55	2.40	0.09	3.19	5.97	8.83
Downside risk (5% VaR)	146,085	2.35	1.38	4.02	0.00	0.00	2.79	8.37
Illiquidity (IILIQ)	146,085	1.07	0.21	3.74	-0.02	0.05	0.68	4.20
Bond market beta (β^{Bond})	146,085	0.42	0.31	0.56	0.00	0.00	0.67	1.27

Panel B: Average cross-sectional correlations

	Maturity	Size	DUR	VaR5	ILLIQ	β^{Bond}
Rating	−0.26	−0.18	−0.15	0.25	0.07	0.01
Maturity	1.00	0.06	0.56	−0.04	0.04	0.07
Size		1.00	0.08	0.00	−0.08	0.13
DUR			1.00	−0.09	−0.02	0.11
VaR5				1.00	0.19	0.61
ILLIQ					1.00	0.04

Table 2: Predicting corporate bond returns with bond characteristics

Panel A of this table reports out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of corporate bonds using the 43 bond characteristics, following equation (12) as $f_1(XB)$. The results are presented at the firm-level by constructing value-weighted firm-level bond returns, as well as the firm-level value-weighted bond characteristics, using amount outstanding as weights. The models include OLS with all bond characteristics (OLS), principal component analysis (PCA), partial least square (PLS), LASSO, ridge regression (Ridge), elastic net (ENet), random forest (RF), feedforward neural network (FFN), long short-term memory neural network (LSTM), and forecast combination (Combination). The R^2_{OS} pools prediction errors across firms and over time into a grand panel-level assessment of each model and is defined as,

$$R^2_{OS} = 1 - \frac{\sum_{(it)\in\mathcal{T}_3} \left(r_{it+1} - \hat{r}_{it+1}\right)^2}{\sum_{(it)\in\mathcal{T}_3} r^2_{it+1}}.$$

p-values associated with R^2_{OS} are reported using one-sided test. The full sample covers the periods from July 2002 to December 2017 and is divided into three disjoint time periods i) the training subsample (the first three years, $\mathcal{T}_1$) to estimate the model, ii) the validation subsample (the following two years, $\mathcal{T}_2$) to tune the hyperparameters, and iii) the test subsample (the rest of the sample, $\mathcal{T}_3$) used to evaluate a model's predictive performance. All of the R^2_{OS} associated with machine learning models from column (2) to column (10) are statistically significant with p-values less than 1%. Panel B reports pairwise Diebold-Mariano test statistics comparing the out-of-sample firm-level bond return prediction performance (R^2_{OS}) among the models used in Table 2. Positive numbers indicate the column model outperforms the row model. Numbers in bold denote statistical significance at the 5% level or better.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
	Panel A: Out-of-sample R^2_{OS}									
R^2_{OS}	−3.36	2.07	2.03	1.85	1.89	1.87	2.19	2.37	2.28	2.09
	Panel B: Comparison of monthly out-of-sample prediction using Diebold-Mariano tests									
OLS		**3.07**	**2.89**	**3.45**	**3.53**	**3.59**	**3.82**	**3.85**	**3.28**	**3.38**
PCA			1.14	−1.32	−1.26	−1.40	**2.10**	1.78	0.28	1.85
PLS				−0.79	−0.57	−0.65	1.78	1.70	0.13	1.14
LASSO					0.44	0.40	1.60	1.18	0.86	**2.05**
Ridge						0.15	1.78	1.96	0.86	**2.00**
Enet							1.81	1.08	0.86	**2.10**
RF								1.10	1.74	1.91
FFN									−0.80	1.20
LSTM										1.15

Table 3: Performance of machine learning bond portfolios using corporate bond characteristics

This table reports the monthly performance of value-weighted decile portfolios sorted on out-of-sample machine learning return forecasts using the 43 bond characteristics (i.e., $\hat{r}_{it+1}$ where $(it) \in \mathcal{T}_3$, the test subsample). At the end of each month, we calculate one-month-ahead out-of-sample firm-level bond return predictions for each method, where the firm-level bond returns are value-weighted using amount outstanding as weights. We then sort firms into deciles based on each model's forecasts and construct the value-weighted portfolio (e.g., using the sum of all bonds amount outstanding within the firm as weights) based on the out-of-sample forecasts. Low corresponds to the portfolio with the lowest expected return (decile 1), High corresponds to the portfolio with the highest expected return (decile 10), and High−Low corresponds to the long short portfolio that buys the highest expected return bonds (decile 10) and sells the lowest (decile 1). The returns are in monthly percentage and Newey-West t-statistics are reported in the last column.

	Low	2	3	4	5	6	7	8	9	High	High−Low	
OLS	0.60	0.67	0.62	0.61	0.56	0.59	0.63	0.60	0.53	0.76	0.16	(1.38)
PCA	0.68	0.61	0.68	0.64	0.65	0.70	0.63	0.67	0.65	1.19	0.51	(2.51)
PLS	0.51	0.55	0.57	0.55	0.57	0.58	0.67	0.65	0.68	1.14	0.63	(2.86)
LASSO	0.57	0.50	0.47	0.40	0.42	0.42	0.46	0.59	0.58	0.96	0.39	(2.54)
Ridge	0.58	0.53	0.46	0.46	0.52	0.45	0.62	0.60	0.67	0.91	0.33	(2.15)
Enet	0.54	0.52	0.48	0.35	0.43	0.41	0.45	0.58	0.55	0.97	0.43	(2.67)
RF	0.57	0.69	0.54	0.51	0.52	0.50	0.59	0.55	0.49	1.37	0.79	(2.78)
FFN	0.61	0.63	0.48	0.55	0.49	0.59	0.50	0.59	0.56	1.36	0.75	(2.61)
LSTM	0.53	0.64	0.60	0.53	0.47	0.55	0.56	0.62	0.58	1.32	0.79	(3.33)
Combination	0.71	0.63	0.58	0.50	0.52	0.60	0.65	0.61	0.59	1.38	0.67	(3.41)

Table 4: Predicting corporate bond returns with stock characteristics

Panel A of this table reports out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of corporate bonds using the 94 stock characteristics, following equation (12) as $f_1(XS)$. The results are presented at the firm-level by constructing value-weighted firm-level bond returns, as well as the firm-level value-weighted bond characteristics, using amount outstanding as weights. The models include OLS with all variables (OLS), principal component analysis (PCA), partial least square (PLS), LASSO, Ridge regression (Ridge), Elastic Net (ENet), Random Forest (RF), feed forward neural network (FFN), long short-term memory neural network (LSTM), and forecast combination (Combination). The R^2_{OS} pools prediction errors across firms and over time into a grand panel-level assessment of each model and is defined as,

$$R^2_{OS} = 1 - \frac{\sum_{(it)\in\mathcal{T}_3} \left(r_{it+1} - \hat{r}_{it+1}\right)^2}{\sum_{(it)\in\mathcal{T}_3} r^2_{it+1}}.$$

The full sample covers the periods from July 2002 to December 2017 and is divided into three disjoint time periods i) the training subsample (the first three years, $\mathcal{T}_1$) to estimate the model, ii) the validation subsample (the following two years, $\mathcal{T}_2$) to tune the hyperparameters, and iii) the test subsample (the rest of the sample, $\mathcal{T}_3$) used to evaluate a model's predictive performance. All of the R^2_{OS} associated with machine learning models from column (2) to column (10) are statistically significant with p-values less than 1%. Panel B reports the monthly performance of value-weighted bond portfolios (i.e., High−Low return) sorted on out-of-sample machine learning return forecasts.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
	Panel A: R^2_{OS} using stock characteristics									
Using $f_1(XS)$	−3.09	1.70	1.71	1.61	1.57	1.62	1.80	1.88	2.00	2.02
	Panel B: Performance of machine learning High−Low bond portfolio using stock characteristics									
Using $f_1(XS)$	0.02	0.36	0.43	0.24	0.26	0.24	0.43	0.48	0.52	0.52
	(0.12)	(2.35)	(2.67)	(2.12)	(2.11)	(2.03)	(2.28)	(2.25)	(3.09)	(3.13)

Table 5: Predicting corporate bond returns with bond and stock characteristics

Panel A of this table reports out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of corporate bonds using the combined 137 stock and bond characteristics, following equation (12) as $f_1(XB, XS)$. The results are presented at the firm-level by constructing value-weighted firm-level bond returns, as well as the firm-level value-weighted bond characteristics, using amount outstanding as weights. The models include OLS with all variables (OLS), principal component analysis (PCA), partial least square (PLS), LASSO, Ridge regression (Ridge), Elastic Net (ENet), Random Forest (RF), feed forward neural network (FFN), long short-term memory neural network (LSTM), and forecast combination (Combination). The R^2_{OS} pools prediction errors across firms and over time into a grand panel-level assessment of each model and is defined as,

$$R^2_{OS} = 1 - \frac{\sum_{(it)\in\mathcal{T}_3} \left(r_{it+1} - \hat{r}_{it+1}\right)^2}{\sum_{(it)\in\mathcal{T}_3} r^2_{it+1}}.$$

The full sample covers the periods from July 2002 to December 2017 and is divided into three disjoint time periods i) the training subsample (the first three years, $\mathcal{T}_1$) to estimate the model, ii) the validation subsample (the following two years, $\mathcal{T}_2$) to tune the hyperparameters, and iii) the test subsample (the rest of the sample, $\mathcal{T}_3$) used to evaluate a model's predictive performance. All of the R^2_{OS} associated with machine learning models from column (2) to column (10) are statistically significant with p-values less than 1%. Panel B reports the monthly performance of value-weighted bond portfolios (i.e., High−Low return) formed using both stock and characteristics (XS + XB) versus using only stock characteristics (XS) or bond characteristics (XB).

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
	Panel A: R^2_{OS} using stock and bond characteristics									
Using $f_1(XB, XS)$	−5.38	1.74	1.70	1.62	1.60	1.66	1.89	1.97	2.11	2.09
	Panel B: Comparing machine learning High−Low bond portfolio									
Using $f_1(XB, XS)$	0.11	0.51	0.57	0.41	0.37	0.44	0.68	0.64	0.71	0.65
	(1.18)	(2.45)	(2.35)	(2.15)	(2.13)	(2.25)	(3.13)	(3.11)	(3.19)	(3.08)
Using $f_1(XB, XS)$ − Using $f_1(XB)$	−0.05	0.00	−0.06	0.02	0.04	0.01	−0.11	−0.11	−0.08	−0.02
	(−0.97)	(0.02)	(−1.01)	(0.22)	(0.99)	(0.68)	(−1.26)	(−1.35)	(−1.45)	(−0.92)
Using $f_1(XB, XS)$ − Using $f_1(XS)$	0.09	0.15	0.14	0.18	0.11	0.21	0.25	0.16	0.19	0.13
	(2.33)	(1.81)	(1.88)	(1.78)	(1.38)	(1.77)	(2.86)	(2.22)	(2.00)	(2.15)

Table 6: Predicting corporate bond returns with regression-based hedge ratios

Panel A of this table reports out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of corporate bonds, based on equation (21). Specifically, we generate bond return forecasts, $f_2(XB, XS, \hat{h})$, as a function of stock and bond characteristics, as well as the regression-based hedge ratios ($\hat{h}$). Panel B of the table compares the forecasted bond returns with hedging ratio, $f_2(XB, XS, \hat{h})$, to the bond return forecast obtained using bond characteristics, $f_1(XB)$ (Table 2), or the combined stock and bond characteristics, $f_1(XB, XS)$ (Table 5), based on the Diebold-Mariano test statistics. The results are presented at the firm-level by constructing value-weighted firm-level bond returns, using amount outstanding as weights. The R^2_{OS} pools prediction errors across firms and over time into a grand panel-level assessment of each model and is defined as,

$$R^2_{OS} = 1 - \frac{\sum_{(it)\in\mathcal{T}_3} \left(r_{it+1} - \hat{r}_{it+1}\right)^2}{\sum_{(it)\in\mathcal{T}_3} r^2_{it+1}}.$$

The full sample covers the periods from July 2002 to December 2017 and is divided into three disjoint time periods i) the training subsample (the first three years, $\mathcal{T}_1$) to estimate the model, ii) the validation subsample (the following two years, $\mathcal{T}_2$) to tune the hyperparameters, and iii) the test subsample (the rest of the sample, $\mathcal{T}_3$) used to evaluate a model's predictive performance. All of the R^2_{OS} associated with machine learning models in Panel A from column (2) to column (10) are statistically significant with p-values less than 1%. Numbers in bold in Panel B denote statistical significance at the 5% level or better.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
	Panel A: R^2_{OS}									
Using $f_2(XB, XS, \hat{h})$	−4.37	2.28	2.88	1.93	1.95	1.95	3.05	3.11	4.89	4.95
	Panel B: Comparison of monthly out-of-sample prediction using Diebold-Mariano tests									
Using $f_2(XB, XS, \hat{h})$ − Using $f_1(XB)$	−1.01	**0.21**	**0.85**	0.08	0.06	0.08	**0.86**	**0.74**	**2.61**	**2.86**
Using $f_2(XB, XS, \hat{h})$ − Using $f_1(XB, XS)$	1.01	**0.54**	**1.18**	**0.31**	**0.35**	**0.29**	**1.16**	**1.14**	**2.78**	**2.86**

Table 7: Performance of machine learning bond portfolios using regression-based hedge ratios

This table reports the monthly performance of value-weighted bond portfolios (i.e., High−Low return) formed using regression-based hedge ratios based on equation (21), $f_2(XB, XS, \hat{h})$, versus using bond characteristics only, $f_1(XB)$, or combined stock and bond characteristics, $f_1(XB, XS)$. Numbers in bold denote statistical significance at the 5% level or better.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
Using $f_2(XB, XS, \hat{h})$	0.18	**0.64**	**0.69**	**0.55**	**0.57**	**0.57**	**0.86**	**0.89**	**0.92**	**0.84**
	(1.07)	(2.16)	(2.33)	(2.60)	(2.75)	(2.77)	(3.01)	(2.68)	(2.69)	(3.27)
Using $f_2(XB, XS, \hat{h})$	0.02	**0.13**	0.06	**0.16**	**0.24**	**0.14**	**0.07**	**0.14**	**0.13**	**0.17**
− Using $f_1(XB)$	(0.35)	(2.35)	(1.93)	(2.27)	(2.45)	(2.36)	(2.04)	(2.54)	(2.38)	(2.81)
Using $f_2(XB, XS, \hat{h})$	0.07	**0.13**	**0.12**	**0.14**	**0.20**	**0.13**	**0.18**	**0.25**	**0.21**	**0.19**
− Using $f_1(XB, XS)$	(0.76)	(2.44)	(2.87)	(2.15)	(2.43)	(2.22)	(2.36)	(2.77)	(2.63)	(2.60)

Table 8: Predicting corporate bond returns with machine learning-based hedge ratios

Panel A of this table reports out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of corporate bonds, based on equation (24). Specifically, we generate bond return forecasts, $f_3(XB, XS, \hat{h}(XB))$, as a function of stock and bond characteristics, as well as machine-learning-based hedge ratios, $\hat{h}(XB)$. Panel B of the table compares the forecasted bond returns with machine-learning-based hedge ratios, $f_3(XB, XS, \hat{h}(XB))$, to the bond return forecast obtained using bond characteristics, $f_1(XB)$ (Table 2), or the combined stock and bond characteristics, $f_1(XB, XS)$ (Table 5), or using regression-based hedge ratios, $f_2(XB, XS, \hat{h})$ (Table 6), based on the Diebold-Mariano test statistics. The results are presented at the firm-level by constructing value-weighted firm-level bond returns, using amount outstanding as weights. The R^2_{OS} pools prediction errors across firms and over time into a grand panel-level assessment of each model and is defined as,

$$R^2_{OS} = 1 - \frac{\sum_{(it)\in\mathcal{T}_3} \left(r_{it+1} - \hat{r}_{it+1}\right)^2}{\sum_{(it)\in\mathcal{T}_3} r^2_{it+1}}.$$

The full sample covers the periods from July 2002 to December 2017 and is divided into three disjoint time periods i) the training subsample (the first three years, $\mathcal{T}_1$) to estimate the model, ii) the validation subsample (the following two years, $\mathcal{T}_2$) to tune the hyperparameters, and iii) the test subsample (the rest of the sample, $\mathcal{T}_3$) used to evaluate a model's predictive performance. All of the R^2_{OS} associated with machine learning models in Panel A from column (2) to column (10) are statistically significant with p-values less than 1%. Numbers in bold denote statistical significance at the 5% level or better.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
	Panel A: R^2_{OS}									
Using $f_3(XB, XS, \hat{h}(XB))$	−4.59	2.35	3.07	2.04	2.05	2.05	3.30	3.53	5.67	5.70
	Panel B: Comparison of out-of-sample prediction using Diebold-Mariano tests									
Using $f_3(XB, XS, \hat{h}(XB))$ − Using $f_1(XB)$	−1.23	**0.28**	**1.04**	**0.19**	**0.16**	**0.18**	**1.11**	**1.14**	**3.39**	**3.61**
Using $f_3(XB, XS, \hat{h}(XB))$ − Using $f_1(XB, XS)$	0.79	**0.61**	**1.37**	**0.42**	**0.45**	**0.39**	**1.41**	**1.56**	**3.56**	**3.61**
Using $f_3(XB, XS, \hat{h}(XB))$ − Using $f_2(XB, XS, \hat{h})$	−0.22	**0.07**	**0.19**	**0.11**	**0.10**	**0.10**	**0.25**	**0.42**	**0.78**	**0.75**

Table 9: Performance of machine learning bond portfolios using machine learning-based hedge ratios

This table reports the monthly performance of value-weighted bond portfolios (i.e., High−Low return) formed using machine-learning-based hedge ratios based on equation (24), $f_3(XB, XS, \hat{h}(XB))$, versus using bond characteristics only, $f_1(XB)$, or combined stock and bond characteristics, $f_1(XB, XS)$, or using regression-based hedge ratios, $f_2(XB, XS, \hat{h})$. Numbers in bold denote statistical significance at the 5% level or better.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
Using $f_3(XB, XS, \hat{h}(XB))$	0.16	**0.65**	**0.71**	**0.54**	**0.57**	**0.58**	**0.89**	**0.93**	**1.00**	**0.89**
	(0.53)	(2.61)	(2.49)	(2.49)	(2.52)	(2.47)	(2.71)	(2.84)	(3.22)	(4.68)
Using $f_3(XB, XS, \hat{h}(XB))$	0.00	**0.14**	0.08	**0.15**	**0.24**	**0.15**	**0.10**	**0.18**	**0.21**	**0.22**
− Using $f_1(XB)$	(0.02)	(2.14)	(1.81)	(2.43)	(2.55)	(2.61)	(2.05)	(2.07)	(2.12)	(2.41)
Using $f_3(XB, XS, \hat{h}(XB))$	0.05	**0.14**	**0.14**	**0.13**	**0.20**	**0.14**	**0.21**	**0.33**	**0.29**	**0.24**
− Using $f_1(XB, XS)$	(0.76)	(2.25)	(2.41)	(2.42)	(2.56)	(2.41)	(2.21)	(2.44)	(2.51)	(2.75)
Using $f_3(XB, XS, \hat{h}(XB))$	−0.02	0.01	0.02	−0.01	0.00	0.01	0.03	0.04	0.08	0.05
− Using $f_2(XB, XS, \hat{h})$	(−0.22)	(0.21)	(0.44)	(−0.10)	(0.02)	(0.10)	(0.43)	(0.54)	(1.15)	(0.79)

Table 10: Comparison of hedge ratios

This table provides comparison of different hedge ratios based on the mean square errors (MSEs), defined as the the average sum of squared differences between the regression-based or machine-learning-based hedge ratios and the benchmark hedge ratio. The hedge ratios include (1) the regression-based hedge ratios in Section 5, and (2) the machine learning-based hedge ratios in Section 6. The benchmark hedge ratio used to calculate MSEs is based on equation (3) in Section 2.1. Panel B reports the MSEs for the subsample based on the firm-level credit rating of individual bonds and Panel C reports the MSEs for the subsample based on the firm-level time-to-maturity of individual bonds.

	Regression-	Machine learning-based									
	based	OLS	PCA	PLS	LASSO	Ridge	Enet	RF	FFN	LSTM	Combination
All	0.051	0.097	0.053	0.054	0.053	0.053	0.054	0.055	0.057	0.056	0.054
Investment-grade (Rating $\leq$ 10)	0.033	0.064	0.032	0.031	0.033	0.032	0.032	0.031	0.031	0.031	0.031
Non-investment-grade (Rating $>$ 10)	0.018	0.033	0.021	0.023	0.021	0.021	0.021	0.024	0.025	0.025	0.023
Short-maturity (1 $\leq$ Maturity $<$ 3)	0.003	0.007	0.003	0.003	0.004	0.004	0.004	0.004	0.005	0.004	0.004
Medium-maturity (3 $\leq$ Maturity $<$ 7)	0.024	0.049	0.026	0.027	0.026	0.026	0.026	0.026	0.027	0.026	0.026
Long-maturity (Maturity $\geq$ 7)	0.024	0.042	0.024	0.024	0.023	0.023	0.024	0.025	0.025	0.025	0.024

Figure 1: Variable importance by model for corporate bond return prediction

This figure presents the variable importance for the top 10 most influential firm-level bond characteristics in each model for corporate bond returns, using the 43 bond characteristics as the covariates. For each model, we calculate the reduction in R^2_{OS} from setting all values of a given predictor to zero within each training sample, and average these into a single importance measure for each predictor. Variable importance is an average over all training samples.

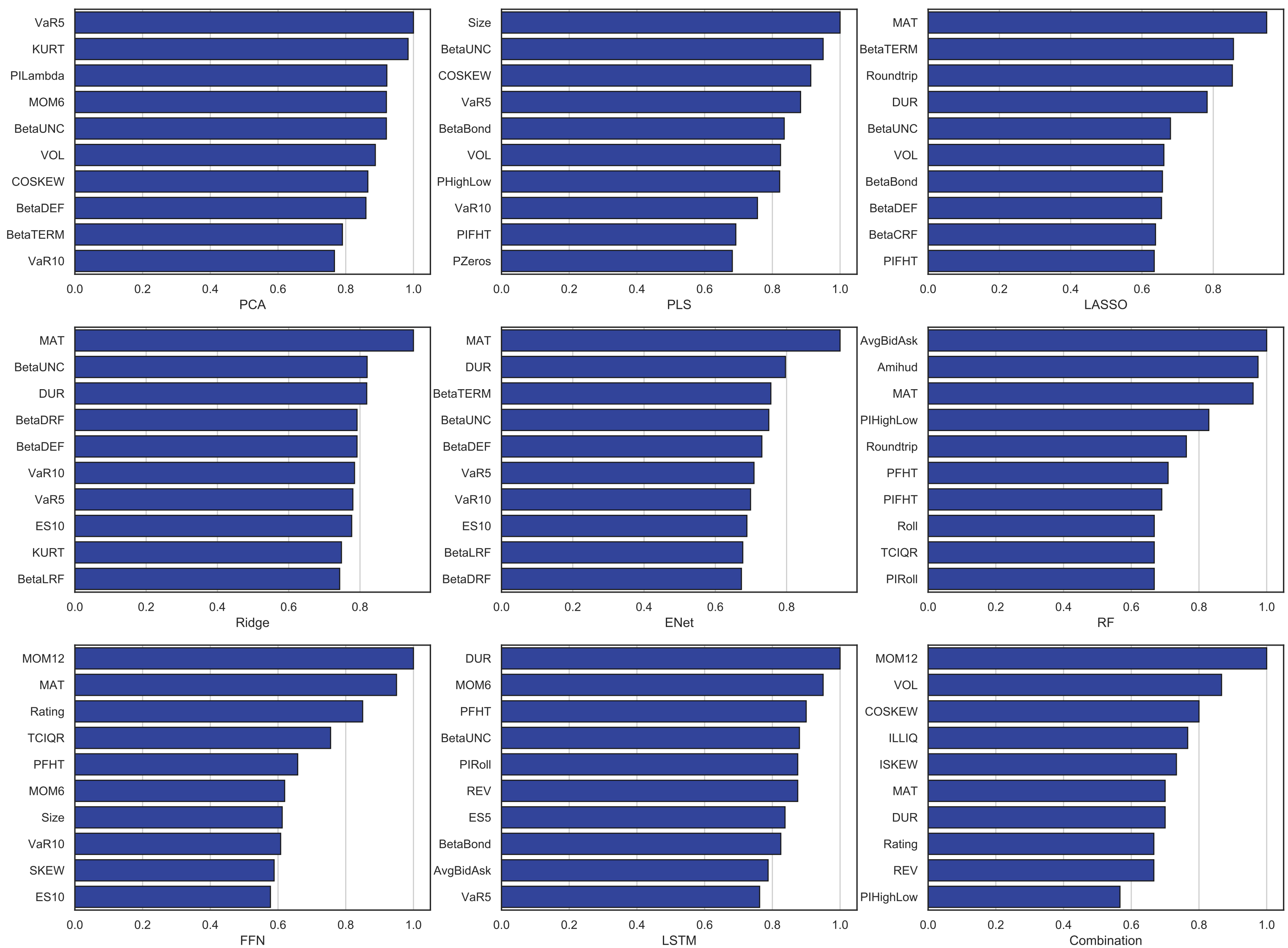


Figure 2: Characteristic importance for corporate bond return prediction

This figure presents the overall rankings of firm-level bond characteristics in each model for corporate bond return prediction. For each model of the nine machine learning methods, we calculate the reduction in R^2_{OS} from setting all values of a given predictor to zero within each training sample, and average these into a single importance measure for each predictor. The importance of each characteristic for each method is ranked and then summed into a single rank. Columns correspond to individual models, and color gradients within each column indicate the most in influential (dark blue) to least in influential (white) variables.

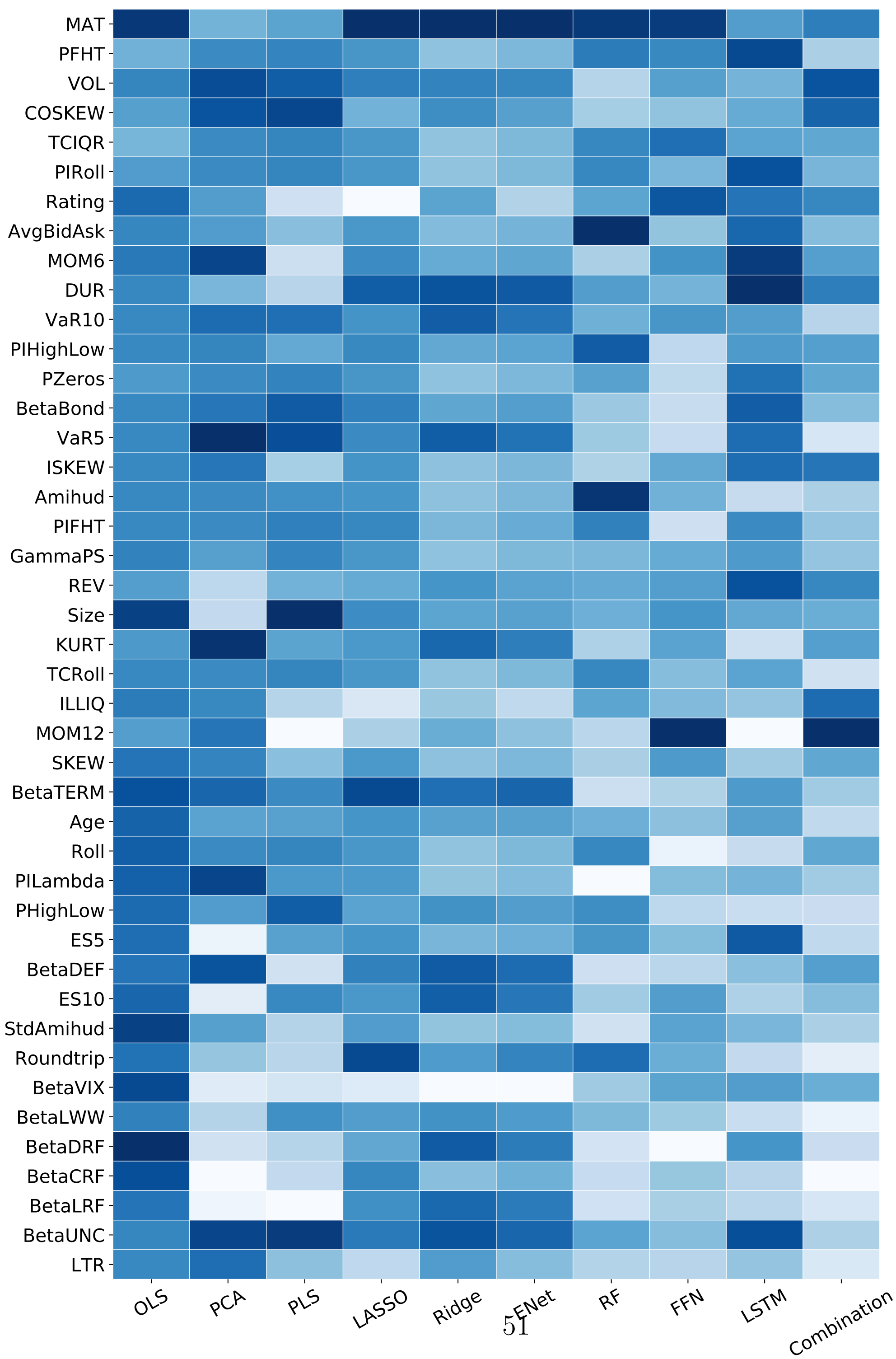
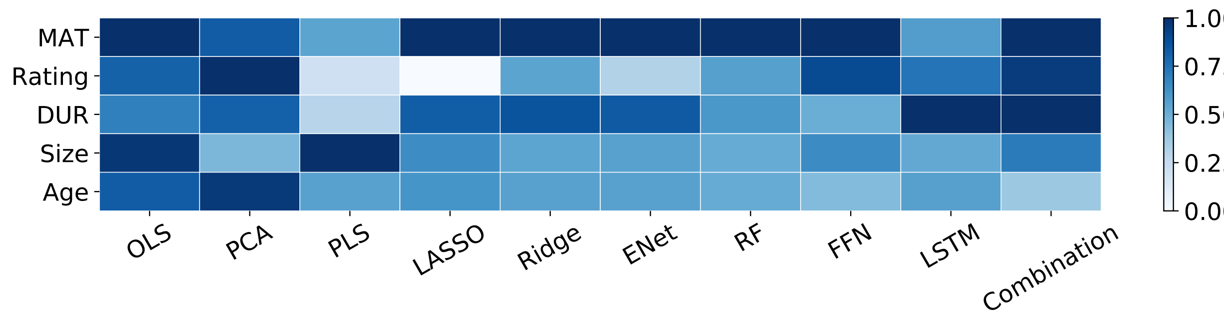


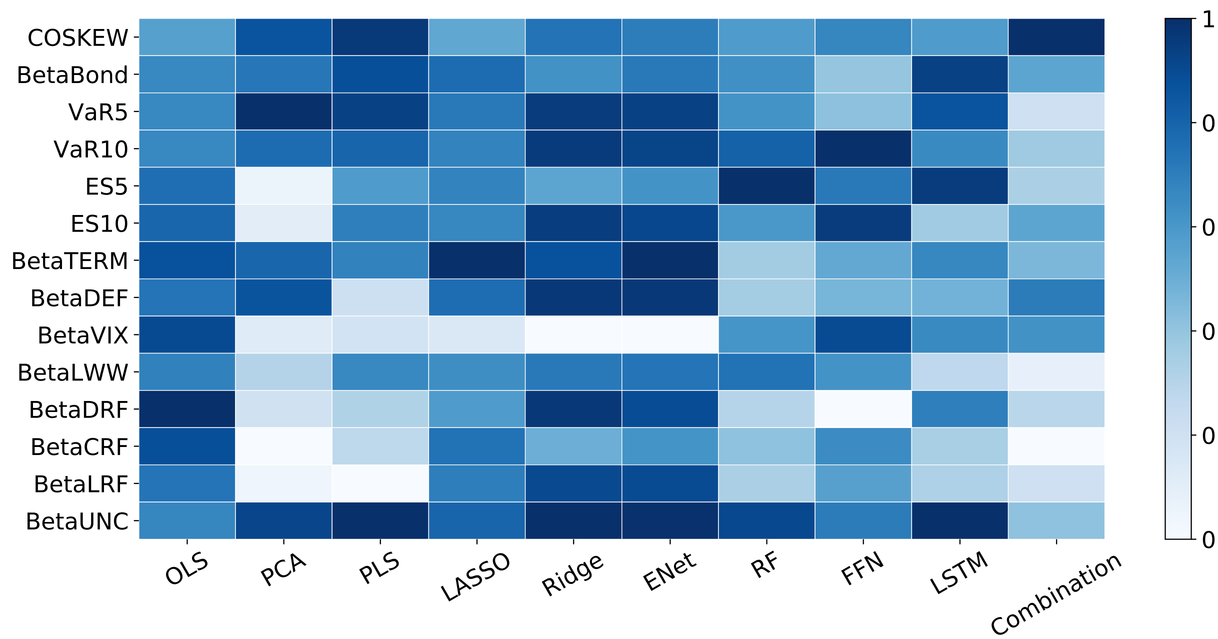
Figure 3: Characteristic importance for corporate bond return prediction based on four characteristic groups

This figure presents the overall rankings of firm-level bond characteristics in each model, within each of the four characteristic groups, for corporate bond return prediction. We classify the 43 corporate bond characteristics into four broad categories (i) bond characteristics related to interest rate risk such as duration and time-to-maturity, (ii) risk measures such as downside risk proxied by Value-at-Risk (VaR) or expected shortfall (ES), total return volatility (VOL), and systematic risk proxied by the bond market beta, default beta, term beta, and macroeconomic uncertainty beta (iii) bond-level illiquidity measures such as average bid and ask price (AvgBidAsk), Amihud and Roll's measures of illiquidity, and (iv) past return characteristics related to bond momentum, short-term reversal, and long-term reversal. For each model of the nine machine learning methods, we calculate the reduction in R^2_{OS} from setting all values of a given predictor to zero within each training sample, and average these into a single importance measure for each predictor. The importance of each characteristic for each method is ranked and then summed into a single rank. Columns correspond to individual models, and color gradients within each column indicate the most in influential (dark blue) to least in influential (white) variables.

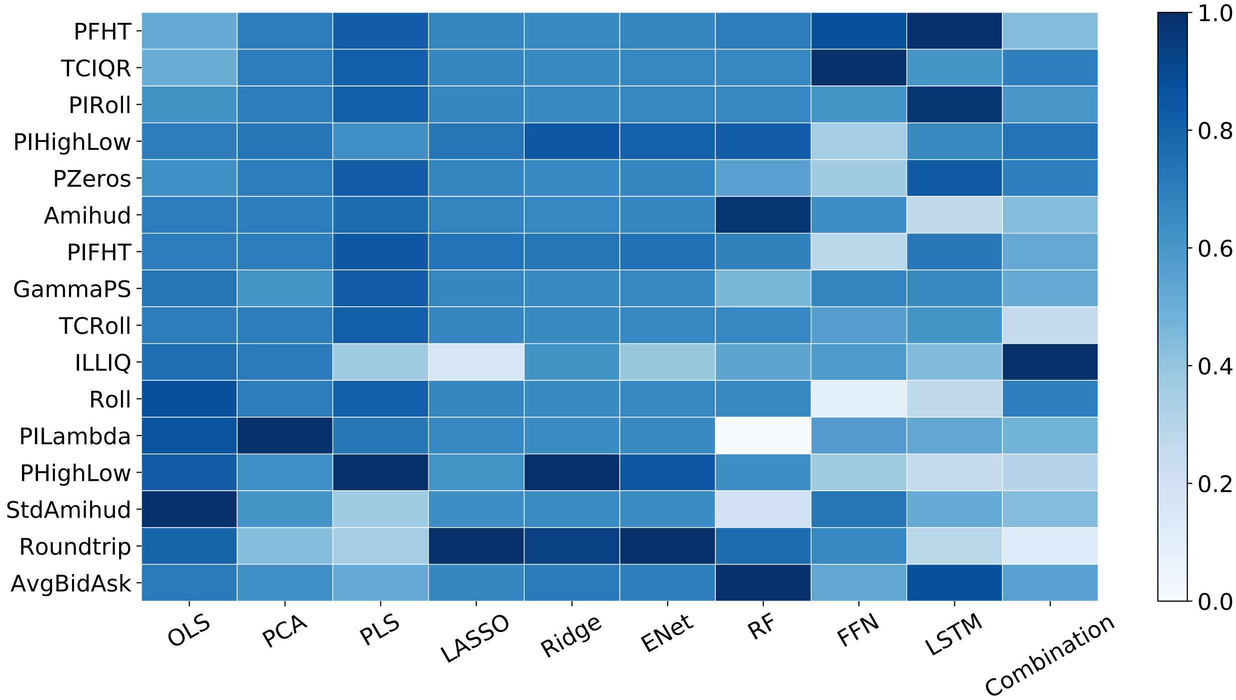
Panel A: Group I characteristics related to interest risk or maturity



Panel B: Group II characteristics related to risk measures



Panel C: Group III characteristics related to illiquidity and illiquidity risk



Panel D: Group IV characteristics related to past return characteristics

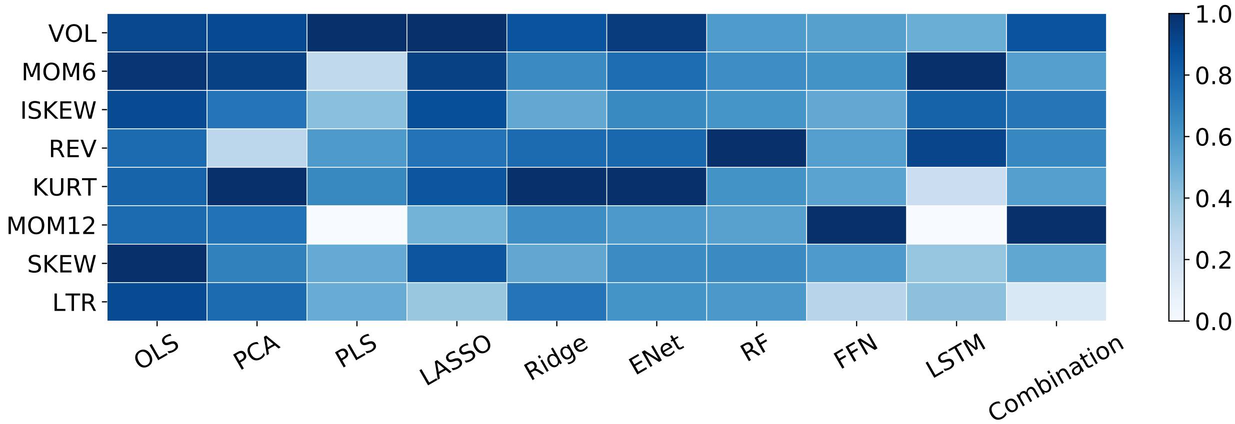


Figure 4: Which characteristic group matter for corporate bond return prediction?

This figure presents the importance of the four characteristic group (Group I to IV), respectively, for all machine learning models. We classify the 43 firm-level corporate bond characteristics into four broad categories (i) bond characteristics related to interest rate risk such as duration and time-to-maturity, (ii) risk measures such as downside risk proxied by Value-at-Risk (VaR) or expected shortfall (ES), total return volatility (VOL), and systematic risk proxied by the bond market beta, default beta, term beta, and macroeconomic uncertainty beta (iii) bond-level illiquidity measures such as average bid and ask price (AvgBidAsk), Amihud and Roll's measures of illiquidity, and (iv) past return characteristics related to bond momentum, short-term reversal, and long-term reversal. The figure reports the sum of the importance of each characteristic for each method, within each characteristic group. Columns correspond to individual models, and color gradients within each column indicate the most in influential (dark blue) to least in influential (white) characteristic group.

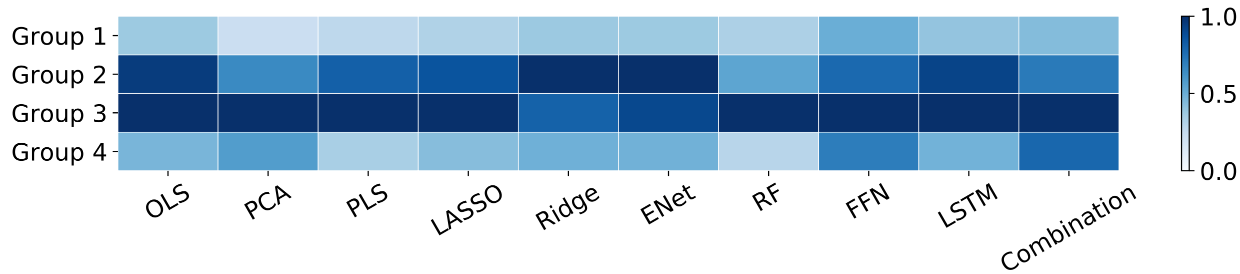
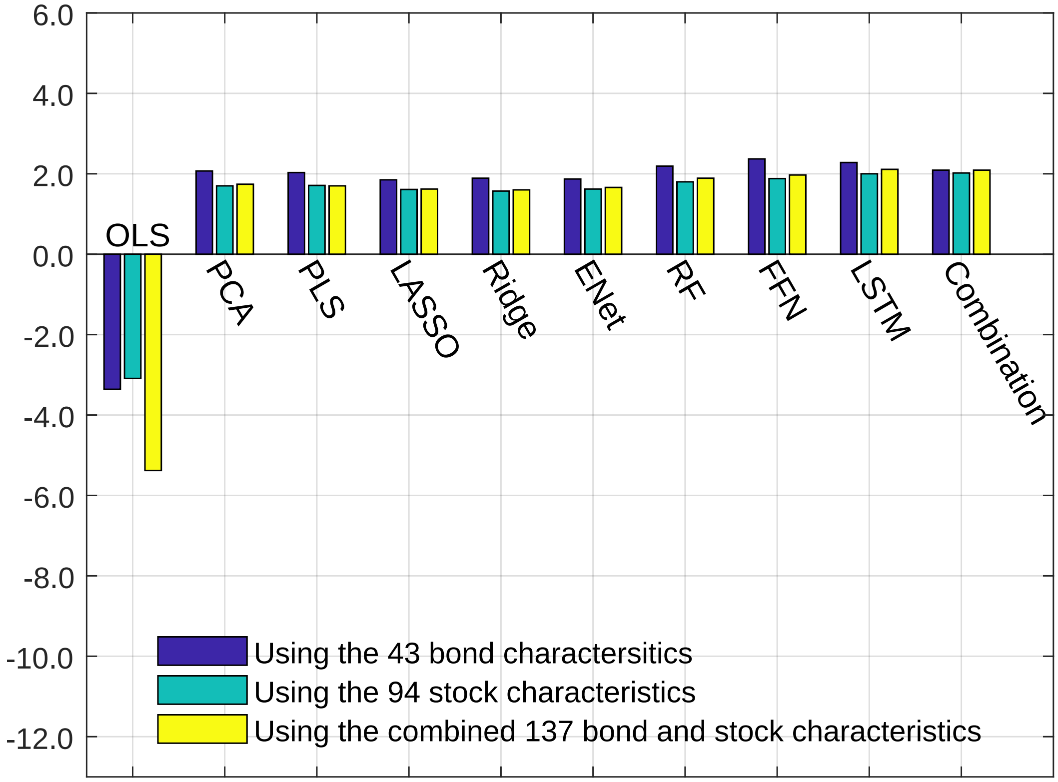


Figure 5: Out-of-Sample R^2_{OS} of Corporate Bond Returns

This figure presents the out-of-sample R-squared (R^2_{OS}, in percentage) of firm-level corporate bond returns using OLS, principal component analysis (PCA), partial least square (PLS), LASSO, ridge regression (Ridge), elastic net (Enet), random forest (RF), feedforward neural network (FFN), long short-term memory neural network (LSTM), and forecast combination (Combination). Figure reports the R^2_{OS} for using 43 bond characteristics only ($f_1(XB)$), 94 stock characteristics only ($f_1(XS)$), and the 137 combined bond and stock characteristics ($f_1(XB, XS)$).



Predicting Corporate Bond Returns: Merton Meets Machine Learning

Online Appendix

Table OA1 provides robust checks of the main results in Table 3 and reports the monthly performance of value-weighted decile portfolios sorted on out-of-sample machine learning return forecasts using the 43 bond characteristics after taking into account of transaction costs.

Table OA2 provides robust checks of the main results in Table 3 and reports the conditional portfolio performance, across different economic states based on the Chicago Fed National Activity Index (CFNAI).

Table OA3 provides robust checks of the main results in Table 3 using maturity-matched Treasury returns to calculate bond excess returns.

Table OA4 reports robust checks of the main results by excluding financial firms with with SIC codes between 6000 and 6999.

Table OA5 shows that the machine learning methods provide strong forecasting power using the stock characteristics, consisting with the findings of Gu, Kelly, and Xiu (2020).

Section OA1 provides details on various machine learning methods.

Table OA1: Performance of machine learning bond portfolios using corporate bond characteristics after transaction costs

This table reports the monthly performance of value-weighted decile portfolios sorted on out-of-sample machine learning return forecasts using the 43 bond characteristics (i.e., $\hat{r}_{it+1}$ where $(it) \in \mathcal{T}_3$, the test subsample), after taking into account transaction costs. At the end of each month, we calculate one-month-ahead out-of-sample firm-level bond return predictions for each method, where the firm-level bond returns are value-weighted using amount outstanding as weights. We then sort firms into deciles based on each model's forecasts and construct the value-weighted portfolio (e.g., using the sum of all bonds amount outstanding within the firm as weights) based on the out-of-sample forecasts. Low corresponds to the portfolio with the lowest expected return (decile 1), High corresponds to the portfolio with the highest expected return (decile 10), and High−Low corresponds to the long short portfolio that buys the highest expected return bonds (decile 10) and sells the lowest (decile 1). The returns are in monthly percentage and Newey-West t-statistics are reported in the last column.

	Low	2	3	4	5	6	7	8	9	High	High−Low	
OLS	0.58	0.65	0.60	0.59	0.54	0.57	0.61	0.58	0.51	0.74	0.13	(1.25)
PCA	0.67	0.60	0.67	0.63	0.64	0.69	0.63	0.66	0.64	1.19	0.49	(2.44)
PLS	0.50	0.54	0.56	0.54	0.55	0.57	0.66	0.64	0.66	1.13	0.61	(2.78)
LASSO	0.55	0.48	0.45	0.38	0.40	0.40	0.44	0.57	0.56	0.94	0.35	(2.32)
Ridge	0.56	0.51	0.44	0.44	0.50	0.43	0.60	0.58	0.65	0.89	0.29	(2.04)
Enet	0.52	0.50	0.47	0.33	0.41	0.39	0.43	0.56	0.54	0.95	0.40	(2.57)
RF	0.56	0.67	0.52	0.49	0.50	0.48	0.57	0.54	0.47	1.35	0.75	(2.63)
FFN	0.60	0.61	0.46	0.53	0.47	0.58	0.48	0.57	0.54	1.34	0.71	(2.55)
LSTM	0.51	0.62	0.58	0.51	0.45	0.53	0.54	0.60	0.56	1.31	0.76	(3.12)
Combination	0.69	0.61	0.56	0.48	0.50	0.58	0.63	0.59	0.58	1.37	0.64	(3.25)

Table OA2: Performance of machine learning bond portfolios in different economic states

This table reports the monthly performance of High− Low value-weighted decile portfolios sorted on out-of-sample machine learning return forecasts using the 43 bond characteristics (i.e., $\hat{r}_{it+1}$ where $(it) \in \mathcal{T}_3$, the test subsample), in different economic states based on the Chicago Fed National Activity Index (CFNAI). At the end of each month, we calculate one-month-ahead out-of-sample firm-level bond return predictions for each method, where the firm-level bond returns are value-weighted using amount outstanding as weights. We then sort firms into deciles based on each model's forecasts and construct the value-weighted portfolio (e.g., using the sum of all bonds amount outstanding within the firm as weights) based on the out-of-sample forecasts. Low corresponds to the portfolio with the lowest expected return (decile 1), High corresponds to the portfolio with the highest expected return (decile 10), and High−Low corresponds to the long short portfolio that buys the highest expected return bonds (decile 10) and sells the lowest (decile 1). The returns are in monthly percentage and Newey-West t-statistics are reported.

	CFNAI > 0 (good economic state)		CFNAI < 0 (bad economic state)	
	Average return	t-stat	Average return	t-stat
OLS	0.14	(1.38)	0.11	(1.32)
PCA	0.58	(2.66)	0.40	(2.51)
PLS	0.70	(2.79)	0.51	(2.56)
LASSO	0.44	(2.34)	0.27	(2.24)
Ridge	0.38	(2.21)	0.20	(2.12)
Enet	0.49	(2.39)	0.30	(2.27)
RF	0.86	(3.52)	0.65	(2.84)
FFN	0.85	(3.64)	0.57	(3.21)
LSTM	0.91	(3.45)	0.60	(3.11)
Combination	0.79	(3.66)	0.48	(3.24)

Table OA3: Performance of machine learning bond portfolios using corporate bond characteristics using maturity-matched Treasury returns

This table reports the monthly performance of value-weighted decile portfolios sorted on out-of-sample machine learning return forecasts using the 43 bond characteristics (i.e., $\hat{r}_{it+1}$ where $(it) \in \mathcal{T}_3$, the test subsample), after adjusting for maturity-matched Treasury returns. At the end of each month, we calculate one-month-ahead out-of-sample firm-level bond return predictions for each method, where the firm-level bond returns are value-weighted using amount outstanding as weights. We then sort firms into deciles based on each model's forecasts and construct the value-weighted portfolio (e.g., using the sum of all bonds amount outstanding within the firm as weights) based on the out-of-sample forecasts. Low corresponds to the portfolio with the lowest expected return (decile 1), High corresponds to the portfolio with the highest expected return (decile 10), and High−Low corresponds to the long short portfolio that buys the highest expected return bonds (decile 10) and sells the lowest (decile 1). The returns are in monthly percentage and Newey-West t-statistics are reported in the last column.

	Low	2	3	4	5	6	7	8	9	High	High−Low	
OLS	0.49	0.56	0.49	0.47	0.42	0.45	0.49	0.44	0.37	0.58	0.09	(0.95)
PCA	0.57	0.50	0.55	0.51	0.51	0.56	0.49	0.51	0.49	1.01	0.44	(2.21)
PLS	0.40	0.44	0.44	0.41	0.43	0.44	0.53	0.49	0.52	0.96	0.56	(2.33)
LASSO	0.46	0.39	0.34	0.27	0.28	0.28	0.32	0.43	0.42	0.78	0.31	(2.24)
Ridge	0.47	0.42	0.33	0.32	0.38	0.31	0.48	0.44	0.51	0.73	0.26	(2.00)
Enet	0.43	0.41	0.35	0.22	0.29	0.27	0.31	0.42	0.39	0.79	0.36	(2.27)
RF	0.46	0.58	0.41	0.37	0.38	0.36	0.45	0.39	0.33	1.19	0.73	(2.43)
FFN	0.50	0.52	0.35	0.42	0.35	0.45	0.36	0.43	0.40	1.17	0.67	(2.41)
LSTM	0.42	0.53	0.47	0.40	0.33	0.41	0.42	0.46	0.42	1.12	0.70	(2.91)
Combination	0.59	0.52	0.45	0.37	0.38	0.46	0.51	0.45	0.43	1.18	0.59	(3.04)

Table OA4: Robustness checks excluding financial firms

This table reports robust checks of the main results by excluding financial firms with with SIC codes between 6000 and 6999. Panel A replicates the main results in Table 2 on the out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of corporate bonds using the 43 bond characteristics, following equation (12) as $f_1(XB)$. Panel B replicates the main results in Table 4, following equation (12) as $f_1(XS)$. Panel C replicates the main results in Table 5, following equation (12) as $f_1(XB, XS)$. Panel D replicates the main results in Table 6, based on equation (21) as $f_2(XB, XS, \hat{h})$. Panel E replicates the main results in Table 7, using regression-based hedge ratios based on equation (21), $f_2(XB, XS, \hat{h})$, versus using bond characteristics only, $f_1(XB)$, or combined stock and bond characteristics, $f_1(XB, XS)$. Panel F replicates the main results in Table 8, based on equation (24). Panel F replicates the main results in Table 9, based on equation (24), $f_3(XB, XS, \hat{h}(XB))$, versus using bond characteristics only, $f_1(XB)$, or combined stock and bond characteristics, $f_1(XB, XS)$, or using regression-based hedge ratios, $f_2(XB, XS, \hat{h})$. Numbers in bold denote statistical significance at the 5% level or better.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
Panel A: Do corporate bond characteristics predict corporate bond returns?										
R^2_{OS}	−3.12	2.07	2.14	2.02	2.01	2.04	2.25	2.30	2.27	2.55
Panel B: Do stock characteristics predict corporate bond returns?										
R^2_{OS}	−3.04	1.73	1.76	1.66	1.60	1.66	1.75	1.99	2.13	2.20
Panel C: Comparing machine learning High−Low bond portfolio										
Using $f_1(XB, XS)$ − Using $f_1(XS)$	0.04	0.13	0.16	0.18	0.11	0.23	0.26	0.36	0.26	0.19
	(1.81)	(2.27)	(1.99)	(1.90)	(1.97)	(2.05)	(3.12)	(3.27)	(2.33)	(2.15)
Using $f_1(XB, XS)$ − Using $f_1(XB)$	−0.14	−0.04	−0.14	−0.13	0.13	−0.15	−0.08	0.02	−0.01	0.21
	(−1.36)	(−0.48)	(−1.23)	(−1.07)	(1.28)	(−0.69)	(−1.04)	(−0.62)	(−1.16)	(−0.41)
Panel D: Predicting corporate bond returns with regression-based hedge ratios										
Using $f_2(XB, XS, \hat{h})$	−4.11	2.15	2.68	1.7	1.75	1.75	3	2.95	4.49	4.74
Panel E: Performance of machine learning High−Low bond portfolio using regression-based hedge ratios										
Using $f_2(XB, XS, \hat{h})$	0.18	**0.59**	**0.63**	**0.52**	**0.54**	**0.55**	**0.80**	**0.83**	**0.88**	**0.83**
	(1.07)	(2.04)	(2.23)	(2.45)	(2.49)	(2.47)	(2.91)	(2.55)	(2.61)	(3.10)
Using $f_2(XB, XS, \hat{h})$ − Using $f_1(XB)$	0.02	**0.12**	**0.07**	**0.15**	**0.25**	**0.15**	**0.08**	**0.16**	**0.14**	**0.16**
	(0.33)	(2.35)	(1.96)	(2.27)	(2.45)	(2.36)	(2.04)	(2.54)	(2.38)	(2.81)
Using $f_2(XB, XS, \hat{h})$ − Using $f_1(XB, XS)$	0.06	**0.12**	**0.11**	**0.14**	**0.19**	**0.14**	**0.17**	**0.23**	**0.25**	**0.20**
	(0.74)	(2.41)	(2.85)	(2.18)	(2.47)	(2.25)	(2.28)	(2.61)	(2.55)	(2.46)
Panel F: Predicting corporate bond returns with machine learning-based hedge ratios										
Using (XS+XB+$\hat{h}(XB)$)	−4.37	2.21	2.94	1.89	1.85	1.90	3.03	3.27	5.33	5.31
Panel G: Performance of machine learning High−Low bond portfolio using machine learning-based hedge ratios										
Using $f_3(XB, XS, \hat{h}(XB))$	0.16	**0.61**	**0.66**	**0.53**	**0.54**	**0.54**	**0.82**	**0.86**	**0.93**	**0.86**
	(0.51)	(2.16)	(2.29)	(2.44)	(2.42)	(2.47)	(2.81)	(2.54)	(2.94)	(4.28)
Using $f_3(XB, XS, \hat{h}(XB))$ − Using $f_1(XB)$	0.00	**0.14**	**0.10**	**0.16**	**0.25**	**0.14**	**0.10**	**0.19**	**0.19**	**0.19**
	(0.02)	(2.23)	(2.01)	(2.31)	(2.45)	(2.52)	(1.94)	(2.21)	(2.44)	(2.65)
Using $f_3(XB, XS, \hat{h}(XB))$ − Using $f_1(XB, XS)$	0.04	**0.14**	**0.14**	**0.15**	**0.19**	**0.13**	**0.19**	**0.26**	**0.30**	**0.23**
	(0.75)	(2.21)	(2.33)	(2.32)	(2.46)	(2.41)	(2.14)	(2.37)	(2.69)	(2.70)
Using $f_3(XB, XS, \hat{h}(XB))$ − Using $f_2(XB, XS, \hat{h})$	−0.02	0.02	0.03	0.01	0.00	−0.01	0.02	0.03	0.05	0.03
	(−0.18)	(0.15)	(0.37)	(−0.14)	(0.02)	(-0.19)	(0.48)	(0.63)	(1.01)	(0.59)

Table OA5: Predicting stock returns with stock characteristics

Panel A of this table reports out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of stocks using the 94 stock characteristics. The models include OLS with all variables (OLS), principal component analysis (PCA), partial least square (PLS), LASSO, Ridge regression (Ridge), Elastic Net (ENet), Random Forest (RF), feedforward neural network (FFN), long short-term memory neural network (LSTM), and forecast combination (Combination). The R^2_{OS} pools prediction errors across firms and over time into a grand panel-level assessment of each model and is defined as,

$$R^2_{OS} = 1 - \frac{\sum_{(it)\in\mathcal{T}_3} (r_{it+1} - \hat{r}_{it+1})^2}{\sum_{(it)\in\mathcal{T}_3} r^2_{it+1}}.$$

The full sample covers the periods from July 2002 to December 2017 and is divided into three disjoint time periods i) the training subsample (the first three years, $\mathcal{T}_1$) to estimate the model, ii) the validation subsample (the following two years, $\mathcal{T}_2$) to tune the hyperparameters, and iii) the test subsample (the rest of the sample, $\mathcal{T}_3$) used to evaluate a model's predictive performance. All of the R^2_{OS} associated with machine learning models from column (2) to column (10) are statistically significant with p-values less than 1%. Panel B reports the monthly performance of value-weighted stock portfolios (i.e., high minus low return) sorted based on out-of-sample machine learning return forecasts using the 94 stock characteristics and 43 bond characteristics.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
Panel A: R^2_{OS} using stock characteristics										
Using XS	−3.83	0.35	0.44	0.12	0.08	0.12	0.49	0.54	0.56	0.60
Panel B: High−Low stock portfolio return										
Using XS	0.48	0.58	0.64	0.48	0.79	0.48	0.89	0.84	1.04	1.10
	(1.38)	(1.95)	(1.92)	(2.33)	(2.61)	(2.33)	(2.74)	(2.92)	(2.32)	(2.79)

OA1 Methodology

The excess return of asset i (either a corporate bond or stock) at time $t+1$, $r_{i,t+1}$, is defined as:

$$r_{i,t+1} = \mathrm{E}_t(r_{i,t+1}) + \varepsilon_{i,t+1}, \tag{OA1}$$

where

$$\mathrm{E}_t(r_{i,t+1}) = g^*(z_{i,t}) \tag{OA2}$$

is the time-t expected return and $g(.)$ is a flexible function of asset i's P-dimensional characteristics, i.e, $z_{i,t} = (z_{i,1,t}, \cdots, z_{i,P,t})'$. For ease of exposition, we assume a balanced panel of assets' returns in this section. We index assets by $i = 1, \cdots, N$ and months by $t = 1, \cdots, T$, where N is the number of assets at time t.

OA1.1 Linear Regression

The linear prediction regression is perhaps the least complex but most widely-used method in the literature, which assumes that $g^*(\cdot)$ can be approximated by a linear function as:

$$g(z_{i,t}; \theta) = z'_{i,t}\theta, \tag{OA3}$$

where $\theta = (\theta_1, \cdots, \theta_P)'$ can be estimated by the ordinary least squares (OLS) via the following optimization problem:

$$\min_{\theta} \mathcal{L}(\theta) \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(r_{i,t+1} - g\left(z_{i,t}; \theta\right)\right)^2. \tag{OA4}$$

The estimate of θ in (OA4) is unbiased and efficient if the number of predictors (P) is relatively small, while T is relatively large. In the real world, unfortunately, P is usually comparable with, or even larger than, T, which raises an overfitting issue and makes the OLS estimate inefficient or even inconsistent. To deal with large P, in the following we introduce nine representative machine learning methods that have been recently used in the finance literature.

OA1.2 Penalized Linear Regression: LASSO, Ridge, and Elastic Net

One of the most commonly used methods for reducing the overfitting issue in (OA4) is to add a penalty term to the objective function. The penalty is imposed to tradeoff between mechanically deteriorating in-sample performance of a model and improving its out-of-sample stability. Instead of (OA4), θ can be estimated via:

$$\min_{\theta} \mathcal{L}(\theta; .) \equiv \mathcal{L}(\theta) + \phi(\theta; .), \tag{OA5}$$

where $\phi(\theta; .)$ is the penalty on θ. Depending on the functional form of $\phi(\theta; .)$, the estimates of some elements of θ can be regularized and shrunk towards zero.

Specifically, in the machine learning literature, a general penalty function is:

$$\phi(\theta;\lambda,\rho)=\lambda(1-\rho)\sum_{j=1}^{P}|\theta_j|+\frac{1}{2}\lambda\rho\sum_{j=1}^{P}\theta_j^2, \tag{OA6}$$

where $\lambda > 0$ is a hyperparameter controlling for the amount of shrinkage; the larger the value of λ, the greater the amount of shrinkage. The estimation model in (OA5) reduces to the standard OLS if $\lambda = 0$. When $\rho = 0$, (OA5) corresponds to LASSO, which sets a subset of θ to exactly zero. In this sense, the LASSO is a sparsity modelling technique and can be used for a variable selection. When $\rho = 1$, (OA5) corresponds to the Ridge regression, which shrinks all coefficient estimates closer to zero but does not impose exact zeros anywhere. Thus, the ridge regression is a dense modeling technique and prevents coefficients from becoming unduly large in magnitude. Finally, if ρ has a value between 0 and 1, we have the "elastic net" penalty, representing a compromise between the Ridge and LASSO. Inheriting from the ridge regression, one advantage of elastic net is that it can handle highly-correlated characteristics (Diebold and Shin, 2019).

OA1.3 Dimension Reduction: PCA and PLS

Based on equations (OA1)–(OA3), the excess return can be rewritten as:

$$r_{i,t+1}=z'_{i,t}\theta+\varepsilon_{i,t+1}. \tag{OA7}$$

With a matrix notation, we have:

$$R=Z\theta+E, \tag{OA8}$$

where R is an $NT\times 1$ vector of $r_{i,t+1}$, Z is an $NT\times P$ matrix of stacked predictors $z_{i,t}$, and E is an $NT\times 1$ vector of residuals $\varepsilon_{i,t+1}$.

Since P is relatively large, dimension reduction is an efficient way to attenuate over-fitting by projecting a large number of characteristics into a small number of factors. Two main dimension reduction techniques are the principal components analysis (PCA) and the partial least squares (PLS). Specifically, PCA requires a transformation of a set of asset characteristics into independent principal components, so that the first one has the largest variance, the second one has the second largest, and so on. Then, it uses a few leading components to represent all the asset characteristics and to predict asset returns. Mathematically, the j^{th} principal component, Zw_j, can be solved as:

$$w_j=\text{argmax}_w\text{Var}\left(Zw\right),\ \text{s.t. } w'w=1,\ \text{Cov}\left(Zw,Zw_l\right)=0,\ l=1,2,\cdots,j-1. \tag{OA9}$$

From (OA9), it is apparent that PCA is to maximize the common variation across all the characteristics and its first K principal components represent the strongest variables that explain the variations of the P characteristics. However, there is no guarantee that they are close to the best set of variables that predicts future asset returns. Indeed, this is not surprising since no information about asset returns is used to find the PCA predictors. In the worst case scenario, if an individual characteristic has the largest variance albeit little ability to predict asset returns, it may well be chosen as the first predictor as long as it is uncorrelated with the other characteristics. Our results indicate that this issue does not occur in our setting with the corporate bond data, but it seems possible with the equity data as suggested by Gu, Kelly, and Xiu (2020).

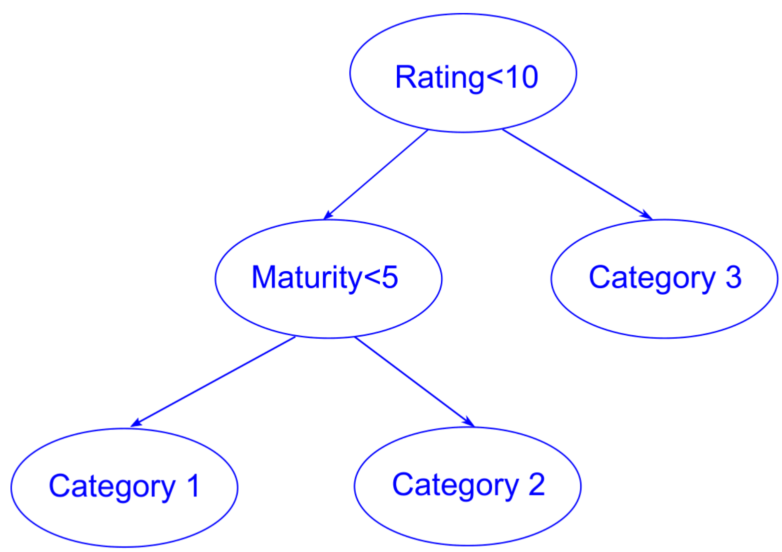
In contrast to PCA, PLS is to link the asset characteristics to the asset returns. In our context, it searches K linear combinations of Z to maximize its covariance with R. The weights w_j used to construct j^{th} PLS component solve for

$$\max_{\nu,w_j} \ \mathrm{Cov}(R\nu, Zw_j), \ \text{s.t. } w_j'w_j = 1, \ \mathrm{Cov}\left(Zw_j, Zw_l\right) = 0, \ l = 1, 2, \cdots, j-1. \tag{OA10}$$

OA1.4 Random Forests

Unlike linear models, random forests are fully nonparametric and have different logic from traditional regressions. A tree "grows" in a sequence of steps. At each step, a new "branch" separates the data based on one of the return predictors. The final outputs are the average values of returns in each partition sliced by predictors. Figure OA1 shows an example with two predictors, "Rating" and "Maturity". The tree separates to a partition based on characteristic values. First, observations are sorted by Ratings. Those above the breakpoint of 10 are assigned to Category 3. Those with lower Ratings (investment bonds) are then further sorted by Maturity. Bonds with lower than one year maturity are assigned to Category 1, while the rest go into Category 2.

Figure OA1: Regression tree example



Mathematically, we can express the expected return approximation function as:

$$g(z_{i,t}; \theta, K, L) = \sum_{k=1}^{K} \theta_k 1_{\{z_{i,t} \in C_k(L)\}}, \tag{OA11}$$

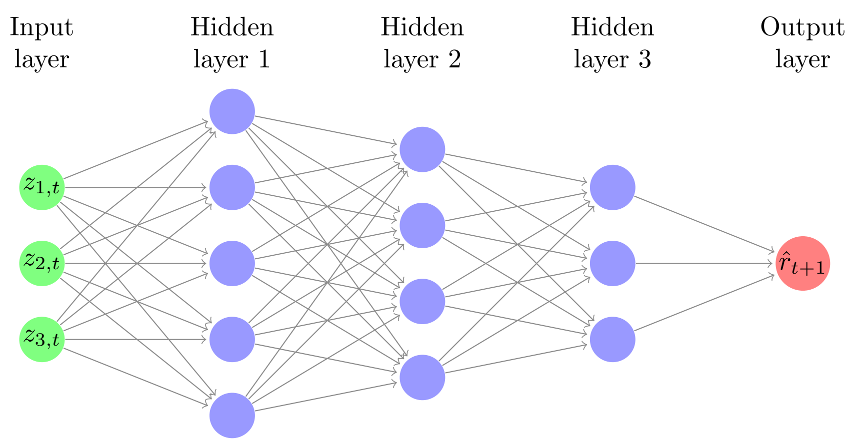
where K is the number of "leaves" (terminal nodes), and L is the depth, $C_k(L)$ is one of the K partitions of the data, $1_{\{\cdot\}}$ is an indicator function, and θ_k is defined to be the sample average of outcomes within the partition. The prediction equation in Figure OA1 is:

$$g\left(z_{i,t}; \theta, 3, 2\right) = \theta_1 1_{\{rating_{i,t}<10\}} 1_{\{maturity<1\}} + \theta_2 1_{\{rating_{i,t}<10\}} 1_{\{maturity\geq 1\}} + \theta_3 1_{\{rating_{i,t}\geq 10\}}. \tag{OA12}$$

To estimate θ in (OA11), we follow the algorithm of Breiman, Friedman, Olshen, and Stone (1984). At each new level, we choose one variable from the set of predictors and the split value to maximize the discrepancy which we call "impurity" among the average outcome returns in each bin. Single the tree is subject to potential over-fitting problems, so it is rarely used without some regulation methods. In our analysis, we build a set of decorrelated trees which are estimated separately and then averaged out as an "ensemble" tree. Such modeling framework is known as "random forests", which is a general procedure known as "bagging" (Breiman, 2001). Through averaging outcomes, random forests can reduce the overfit in individual bootstrap samples and make the predictive performance more stable.

OA1.5 Feed-Forward Neural Network

Figure OA2: Feed-forward neural network with three hidden layers



As a typical neural network, feed-forward neural networks (FFN), as shown in Figure OA2, include an "input layer" of raw predictors, one or more "hidden layers" that interact and nonlinearly transform the predictors, and an "output layer" that aggregates hidden layers into outcome prediction. The information that flows from input layer to hidden layers is transformed and aggregated into an output at the output layer. The model becomes more flexible by adding hidden layers between the inputs and output. Each hidden layer takes the output from the previous layer and transforms it into an output as:

$$z_K^l = g\big(b^{l-1} + (z^{l-1})'W^{l-1}\big), \tag{OA13}$$

where $g(\cdot)$ is the nonlinear "activation function" to its aggregated signal before sending its output to the next layer. The final output is:

$$G(z, b, W) = b^{L-1} + (z^{L-1})'W^{L-1}, \tag{OA14}$$

where $G(\cdot)$ is the linear form of drawing information from the last hidden layer output.

There are many choices for the nonlinear activation functions, and we adopt the most commonly used rectified linear unit (ReLU), which is defined as:

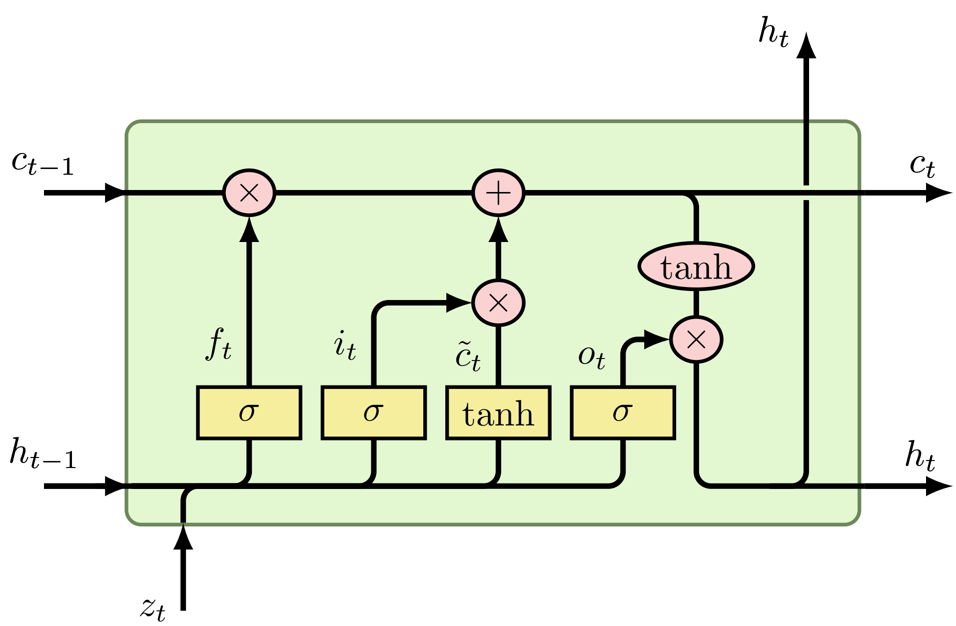
$$\mathrm{ReLU}(z_k) = \max(z_k, 0). \tag{OA15}$$

In this paper, we apply one hidden layer for FFN, with the idea of achieving better performance about shallow learning as in Gu, Kelly, and Xiu (2020).

OA1.6 Long Short-Term Memory Neural Network

In financial markets, many predictors have long-term effects on stock returns. For example, return volatility is known to have a long memory effect (see, e.g., Lo, 1991), and using only one lag of volatility is unlikely to bring volatility to its full potential when predicting future returns. To deal with this type of long-term dependencies, we use a more complex LSTM model (Hochreiter and Schmidhuber, 1997), which transform a sequence of input variables to another output sequence, with the same set of parameters at each step.

Figure OA3: Long Short-Term Memory Networks



Specifically, the LSTM model takes the current input variable z_t and the previous hidden state h_{t-1} and performs a non-linear transformation to get the current state h_t:

$$h_t = g\left(W_h^{(c)} h_{t-1} + W_z^{(c)} z_t + w_0^{(c)}\right). \tag{OA16}$$

This type of structure is powerful if only the immediate past is relevant, but is not suited if the time series dynamics are driven by events that occur in the distant past. We can think of an LSTM as a flexible hidden state space model for a large dimensional system. The LSTM is composed of a cell (the memory part of the LSTM unit) and three "regulators" of the flow of information inside the LSTM unit: an input gate, a forget gate, and an output gate.

The memory cell is created with current input z_t and previous hidden state h_{t-1}:

$$\tilde{c}_t = \tanh\left(W_h^{(c)} h_{t-1} + W_z^{(c)} z_t + w_0^{(c)}\right). \tag{OA17}$$

The input and forget gates control for the memory cell, and the output gate controls for the amount of information stored in the hidden state:

$$\begin{aligned} \text{input}_t &= g\left(W_h^{(i)} h_{t-1} + W_z^{(i)} z_t + w_0^{(i)}\right), \\ \text{forget}_t &= g\left(W_h^{(f)} h_{t-1} + W_z^{(f)} z_t + w_0^{(f)}\right), \\ \text{out}_t &= g\left(W_h^{(o)} h_{t-1} + W_z^{(o)} z_t + w_0^{(o)}\right). \end{aligned} \tag{OA18}$$

Define the element-wise product by $\odot$, the final memory cell and hidden state are given by:

$$\begin{aligned} c_t &= \text{forget}_t \odot c_{t-1} + \text{input}_t \odot \tilde{c}_t \\ h_t &= \text{out}_t \odot \tanh(c_t) \end{aligned} \tag{OA19}$$

Figure OA3 presents the diagram of a long short-term memory network. We consider one hidden layer LSTM method for our model comparison.

OA1.7 Forecast Combination

Let $\hat{r}_{i,t+1}^{(m)}$ be asset i's expected return estimated with method m ($m = 1, \cdots, M$) and $M = 8$ be the number of methods, consisting of PCA, PLS, LASSO, Ridge, Enet, RF, FFN, and LSTM. We equally combine $\hat{r}_{i,t+1}^{(m)}$ to obtain a new prediction of expected return:

$$\hat{r}_{i,t+1} = \frac{1}{M} \sum_{m=1}^{M} \hat{r}_{i,t+1}^{(m)}. \tag{OA20}$$

The idea behind equation (OA20) is that some of the expected return predictions, $\hat{r}_{i,t+1}^{(m)}$, may have a high variance, equally weighting them can reduce the overall variance dramatically, although it may increase the bias to some extent. The literature provides evidence that the forecast combination method works well for return predictability (e.g., Rapach, Strauss, and Zhou, 2010; Chen, Pelger, and Zhu, 2019).

Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch