Which Characteristic Matters for the Expected Corporate Bond Returns? A Machine Learning Approach*

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

Abstract

We provide a comprehensive study of the expected corporate bond returns in light of machine learning methods and data mining bias using a large set of stock and bond characteristics. We examine whether stock or bond characteristics, individually or jointly, drive the expected corporate bond returns. Using either set of characteristics, we find that machine learning methods substantially improve the out-of-sample predictive power for bond returns, compared to the traditional unconstrained linear models. While stock characteristics produce significant explanatory power for bond returns when used alone, their predictive power is only half of that associated with using corporate bond characteristics, and their incremental improvement relative to bond characteristics is small. In addition, both stock and bond characteristics provide strong forecasting power for future *stock* returns when used alone. However, corporate bond characteristics do not offer additional explanatory above and beyond stock characteristics when we combine both sets of characteristics.

This Version: November 2019

Keywords: Machine learning, corporate bond returns

JEL Classification: G10, G11, C13.

*Title subject to change. We thank Kenneth French, Lubos Pastor, and Robert Stambaugh for making a large amount of historical data publicly available in their online data library. All errors remain our responsibility.

†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

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

¶Associate 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

Hundreds of stock characteristics have been presented as statistically significant predictors of the cross-section of stock returns since 1970 (?). Since then, a few studies show that the majority of the predictive power associated with these characteristics are mostly likely an artifact of data snooping, especially when examined out-of-sample (?; ?; ?; ?). Despite the out-of-sample and post-publication decline of a vast majority of stock characteristics (?), recent studies show that machine learning methods are able to generate robust forecasting power to predict stock returns in the cross-section and time series (?).

Despite the proliferation of stock characteristics or factors to explain the cross-section of stock returns, however, far fewer studies are devoted to the expected corporate bond returns. Debt financing forms a significant portion of firms' capital structures, underscoring the need to study the return predictors in corporate bond markets.[1] 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 et al., 2013), downside risk and short-term reversal (?), which exhibit significant explanatory power for the expected bond returns. Other papers investigate whether a few stock characteristics impact the cross-section of corporate bond returns and find mixed evidence on the predictability (Chordia et al., 2017; Choi and Kim, 2016).

In this paper, we provide a comprehensive study of the expected corporate bond returns using a large set of stock and bond characteristics. We first build a comprehensive data library of 41 corporate bond-level characteristics that are motivated by the existing studies on the cross-section of corporate bond returns literature.[2] We then combine them with the 94 stock characteristics used in ? and ?. Our final sample of the 135 stock and bond-level charac-

[1] Graham et al. (2015) indicate that the average debt-to-assets ratio for public companies was as high as 35% in 2010.

[2] This list of a broad set of corporate bond characteristics 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 and short-term reversal, and distributional characteristics such as return volatility. We present a detail list of the bond characteristics in Section 3 and Appendix of the paper as well as the studies that we follow closely to construct these measures.

teristics cover both the equity and debt market, thus provide a wide range of predictors for corporate bond returns. The benefit of using such an exhaustive list is that it avoids selection biases, thereby reducing the data-snooping problem (Lo and MacKinlay, 1990; **?**). Equity and corporate bonds are contingent claims on firms, but also differ in several key features. First and foremost, bondholders are more sensitive to downside risk compared to stockholders (Hong and Sraer, 2013). Since the upside payoffs of corporate bonds are capped, corporate bond payoffs become concave in the investor beliefs about the underlying fundamentals, whereas equity payoffs are linear in investor beliefs regarding fluctuations in the underlying factors (**?**).[3] Hence it is important to examine whether stock or bond characteristics, individually or jointly, drive the expected corporate bond returns. 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 degrees of freedom and condensing redundant variation among a large set of predictors, with an emphasis on variable selection and dimension reduction techniques (**?**).

In light of machine learning methods and the data mining bias, in this article, we seek to answer the following questions using the comprehensive list of the 135 stock and bond-level predictive characteristics: Do corporate bond characteristics and stock characteristics, individually or combined, predict the expected bond returns? Which set of characteristics plays the major role in explaining the expected bond returns? Do stock characteristics improve the predictive power of bond-level characteristics for bond returns? In addition to corporate bond return predictability, we also investigate stock return predictability motivated by the findings of **?**. We further examine the predictive performance of stock and corporate bond characteristics, respectively, in predicting future *stock* returns. We then investigate whether corporate bond characteristics provide marginal improvement on the predictive power over stock characteristics for future stock returns. Although many existing studies have examined whether a large set of stock characteristics predict the cross-section of stock returns (**?**; **?**), to

[3] In addition, the corporate bond market, due to its over-the-counter trading mechanism and other market features, bears higher liquidity risk. Bond market participants are dominated by institutional investors such as insurance companies, pension funds, and mutual funds who are long-term buy-and-hold investors (Source: Financial Accounts of the United States, Release Z1, Table L.21). Thus liquidity in corporate bond market is lower compared to the stock market in which active trading is partially attributable to the existence of individual investors.

the best of our knowledge, far fewer studies studies are devoted to study whether a comprehensive list of corporate bond characteristics provide forecasting power for the expected *stock* returns.[4]

Following a variety of machine learning methods proposed by ?, we compare and evaluate these approaches for the prediction of corporate bond returns based on their out-of-sample predictive performance. 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). On top of these methods, we also add a long-term memory neural network (LSTM) proposed in ? to capture a long memory effect (?). Moreover, we also rely on the forecast combination method (Combination) which averages individual expected return forecasts from the aforementioned sophisticated machine learning models (?; ?).[5] First, we show that traditional unconstrained linear models such as OLS fail to deliver statistically significant out-of-sample forecasting power for the expected corporate bond returns. The OLS model with all 41 bond characteristics produces a negative out-of-sample R-squared (R^2_{OS}), whereas the machine learning models substantially improve the predictive power with monthly R^2_{OS} in the range of 2.45% to 4.40%. In addition, the predictive power of machine learning models extends to both investment-grade and non-investment-grade bonds, with the highest improvements for investment-grade bonds which cover about 75% of the full sample. Using the ? 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 41 bond characteristics. The machine learning bond portfolios are based on the one-month-ahead out-of-sample bond returns, where the high minus low portfolio corresponds to the long short portfolio that buys the highest expected return bonds (decile 10) and sells the lowest (decile 1). We find that all machine learning forecasts generate economically and statistically significant return spreads

[4]This is partly because of the dearth of high-quality corporate bond data required to construct bond returns and a large set of bond characteristics, and the complex features of corporate bonds such as optionality, seniority, changing maturity, and risk exposure to a number of financial and macroeconomic factors.

[5]We describe the machine learning methods used in our analysis in detail in Section 2.

between high and low bond portfolios, in the range of 0.58% to 1.20% per month, compared to the unconstrained OLS model which delivers the smallest monthly return spread of 0.47%.

We proceed to identify corporate bond characteristics that are important for the expected bond returns while simultaneously controlling for the many other predictors. Following the ranking and variable importance approach in **?** and **?**, we discover influential covariates by measuring the reduction in panel prediction 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 performance of each machine learning model. Our results 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 in general (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 related to bond market beta, default and term beta, (iii) bond-level illiquidity measures such as average bid and ask price (AvgBidAsk), and Amihud or Roll's measures of illiquidity, and (iv) bond return characteristics related to bond momentum and short-term reversal.

Second, we examine whether a comprehensive list of stock characteristics provide forecasting power for the expected corporate bond returns. Recent studies often draw from the well of cross-sectional predictors on a few stock characteristics and find mixed evidence of predictability (Chordia et al., 2017; Choi and Kim, 2016). 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 the expected bond returns compared to the OLS model, for all sample of bonds.[6] Interestingly, the predictive performance is higher for non-investment-grade bonds than investment-grade bonds, for all machine learning methods using the 94 stock characteristics. However, the marginal improvement of forecasting performance

[6]The machine learning models using stock characteristics deliver an R^2_{OS} in the range of 0.34% to 2.75% per month, which is about half of the monthly R^2_{OS} associated with using bond characteristics, which ranges from 2.45% to 4.40%. These results indicate that corporate bond characteristics have about twice as much predictive power for bond returns than stock characteristics.

of stock characteristics relative to bond characteristics is small, as the Diebond-Mariano test statistics indicate that there is no difference in the performance of machine learning models when adding stock characteristics to the bond characteristics in forecasting bond returns.

? show that machine learning offers an improved description of expected return relative to traditional methods in forecasting future stock returns. Motivated by the findings of **?**, we examine the predictive performance of the 94 stock and 41 bond characteristics, respectively, in predicting future *stock* returns. Consistent with the findings in **?**, we show that machine learning methods provide strong forecasting power for future stock returns using the 94 stock characteristics. Moreover, we uncover new evidence that when used alone in the machine learning models, corporate bond characteristics perform well in predicting future stock returns, with out-of-sample R-squared ranging from 0.11% to 0.44% per month. However, when comparing the marginal improvement in the predictive performance, we find that corporate bond characteristics do not offer additional explanatory power above and beyond stock characteristics.

This paper proceeds as follows. Section 2 describes the machine learning methods used in the paper and the performance metrics used to assess the predictive performance for individual bond return forecasts. Section 3 describes the data and variables used in our empirical analyses. Section 4 investigates the forecasting performance of a variety of bond and stock characteristics, individually or jointly, for the expected bond returns. Section 5 examines whether stock or bond characteristics are significant predictors for the expected stock returns. Section 6 concludes the paper.

2 Methodology

In this section, we describe a variety of machine learning methods used in our analysis. These candidate models are examined in measuring equity risk premia in **?**, including generalized linear models with penalization, dimension reduction via principal components regression (PCR) and partial least squares (PLS), regression trees, random forests, and neural networks.

Following **?**, we write the excess return of asset i (either a corporate bond or stock) as

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

where

$$\mathrm{E}_t(r_{i,t+1}) = g^*(z_{i,t}) \tag{2}$$

is the expected return at time t and 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 and discuss the missing data issues in Section 3; 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.

2.1 Linear Regression

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

$$g(z_{i,t}; \theta) = z_{i,t}'\theta, \tag{3}$$

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

$$\min_{\theta} \mathcal{L}(\theta) \equiv \frac{1}{NT} \sum_{t=1}^{N} \sum_{t=1}^{T} \left(r_{i,t+1} - g\left(z_{i,t}; \theta\right)\right)^2. \tag{4}$$

Based on **?**, the estimate of θ in (4) is unbiased and efficient if 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.

2.2 Penalized Linear: LASSO, Ridge and Elastic Net

A most common method for reducing the overfitting issue in (4) is to add a penalty term to the objective function. The penalty is imposed to tradeoff between mechanically deteriorating a model's in-sample performance and improving its stability out-of-sample. Instead of (4), θ can be estimated via

$$\min_{\theta} \mathcal{L}(\theta; .) \equiv \mathcal{L}(\theta) + \phi(\theta; .), \tag{5}$$

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

$$\phi(\theta; \lambda, \rho) = \lambda(1-\rho)\sum_{j=1}^{P}|\theta_j| + \frac{1}{2}\lambda\rho\sum_{j=1}^{P}\theta_j^2, \tag{6}$$

where $\lambda > 0$ is a hyperparameter congrolling for the amount of shrinkage; the larger the value of λ, the greater the amount of shrinkage. the estimate in (5) reduces to the OLS estimate if $\lambda = 0$. When $\rho = 0$, (5) 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 variable selection. When $\rho = 1$, (5) corresponds to the Ridge regression, which shrinks all coefficient estimates closer to zero but does not impose exact zeros anywhere. In this sense, 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 ridge regression, one advantage of elastic net is that it can handle highly-correlated characteristics (**?**).

2.3 Dimension Reduction: PCA and PLS

Based on equations (1)–(3), the excess return can be rewritten as

$$r_{i,t+1} = z_{i,t}'\theta + \varepsilon_{i,t+1}. \tag{7}$$

With matrix notations, we have

$$R = Z\theta + E, \tag{8}$$

where R is the $NT \times 1$ vector of $r_{i,t+1}$, Z is the $NT \times P$ 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 principal components analysis (PCA) and 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:

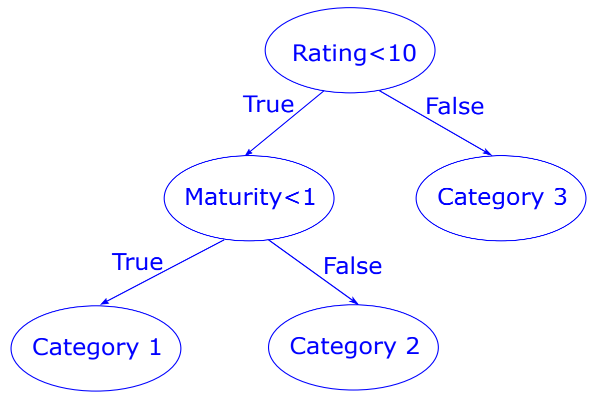
$$w_j = \text{argmax}_w \text{Var}\left(Zw\right), \text{ s.t. } w_j' w_j = 1, \text{ Cov}\left(Zw, Zw_l\right) = 0, \; l = 1, 2, \cdots, j-1. \tag{9}$$

From (9), it is apparent that PCA is to maximize the common variation across all the characteristics and its first K principal components represent the best variables that explain the variations of the P characteristics. However, there is no guarantee that they are close the best variables that predict the future asset returns. Indeed, this is not surprising since no information about asset returns is used in finding the PCA predictors. In the worst case, if an individual characteristic has the largest variance and little ability to predict asset returns, it will be very likely chosen as the first predictor as long as it is uncorrelated with the other characteristics. Of course, this may not happen in the real data like our setting in the bond market, but it seems true in the stock market (**?**).

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

Figure 1. Regression tree example

This figure presents the diagram of a regression tree in the space of two predictors (rating and maturity), based on which, the sample of individual bonds is divided into three categories.



weights used to construct j^{th} PLS component solve for

$$w_j = \text{argmax}_w \ \text{Cov}^2(R, Zw), \ \text{s.t.} \ w'w = 1, \ \text{Cov}\,(Zw, Zw_j) = 0, \ l = 1, 2, \cdots, j-1. \quad (10)$$

2.4 Random Forests

Unlike linear models, random forecasts 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 predictor variables. The final outputs are the average values of returns in each partition sliced by predictors. Figure 1 shows an example with two predictors, "Rating" and "Maturity". The tree separates to a partition based on characteristic values. First, observations are sorted on Rating. Those above the breakpoint of 10 are assigned to Category 3. Those with small Rating values(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.

Mathematically, we can express the expected return approximation function as

$$g(z_{i,t}; \theta, K, L) = \sum_{k=1}^{K} \theta_k 1_{\{z_{i,t} \in C_k(L)\}}, \quad (11)$$

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, $1_{\{\}}$ is an indicator function, and θ_k is defined to be the sample average of outcomes within the partition. The prediction equation in Figure 1 is

$$g\left(z_{i,t};\theta,3,2\right)=\theta_1 1_{\{rating_{i,t}<10\}}1_{\{maturity<1\}}+\theta_2 1_{\{rating_{i,t}<10\}}1_{\{maturity\geq 1\}}+\theta_3 1_{\{rating_{i,t}\geq 10\}}. \quad (12)$$

To estimate θ in (11), we follow the algorithm of **?**. At each new level, we choose one variable from the set of predictors and the split value to maximize the discrepancy which we call it "impurity" among the average outcome returns in each bin. Single tree is facing over fitting problems in prediction, so it's 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 a "ensemble" tree. Such modeling framework is known as "random forests", which is a general procedure known as "bagging" (**?**). Through averaging outcomes, random forests can reduce the overfit in individual bootstrap samples, and make the predictive performance more stable.

2.5 Feed-Forward Neural Network

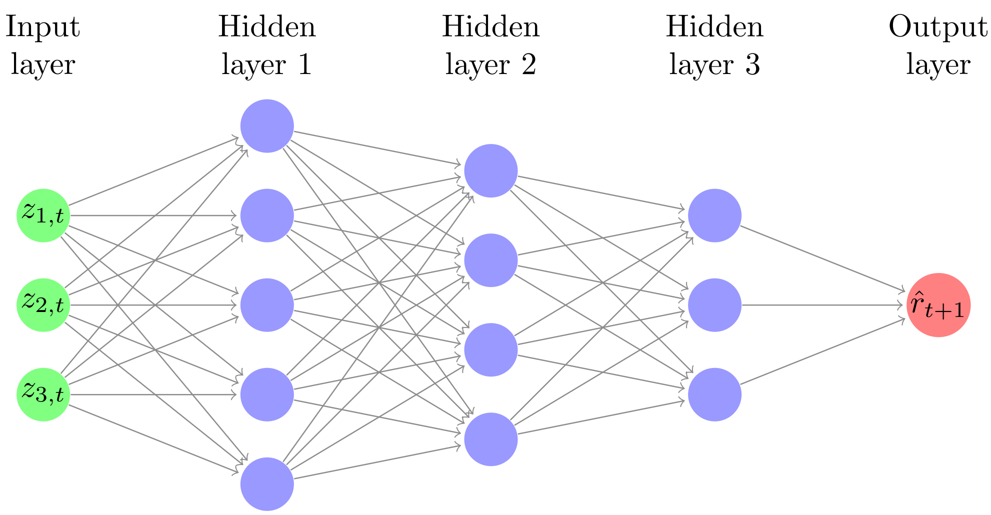
As a typical neural network, feed-forward neural networks (FFN) include an "input layer" of raw predictors, one or more "hidden layers" that interact and nonlinearly transform the predictors, and an "output layer" that aggregates hidden layers into outcome prediction. The information flows from input layer to hidden layers, finally 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:

$$z_K^{(l)}=g(b^{l-1}+z^{(l-1)'}W^{(l-1)}), \quad (13)$$

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

$$G(z,b,W)=b^{L-1}+z^{(L-1)'}W^{(L-1)}, \quad (14)$$

Figure 2. Feed-forward neural network with three hidden layer



where is the linearly form 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:

$$\text{ReLU}(z_k) = \max(z_k, 0). \tag{15}$$

In this article, we apply one hidden layer for FFN, considering better performance about shallow learning from **?**.

2.6 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., **?**), and using only one lag of volatility for predicting returns is unlikely to fully capture the volatility forecasting power. To deal with this type of long-term dependencies, we use a more complex LSTM model (**?**), which transform a sequence of input variables to another output sequence, with the same set of parameters at each step.

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

$$h_t = g\left(W_h^{(c)} h_{t-1} + W_z^{(c)} z_t + w_0^{(c)}\right). \tag{16}$$

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 are further back in the 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}

$$\tilde{c}_t = \tanh\left(W_h^{(c)} h_{t-1} + W_z^{(c)} z_t + w_0^{(c)}\right). \tag{17}$$

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:

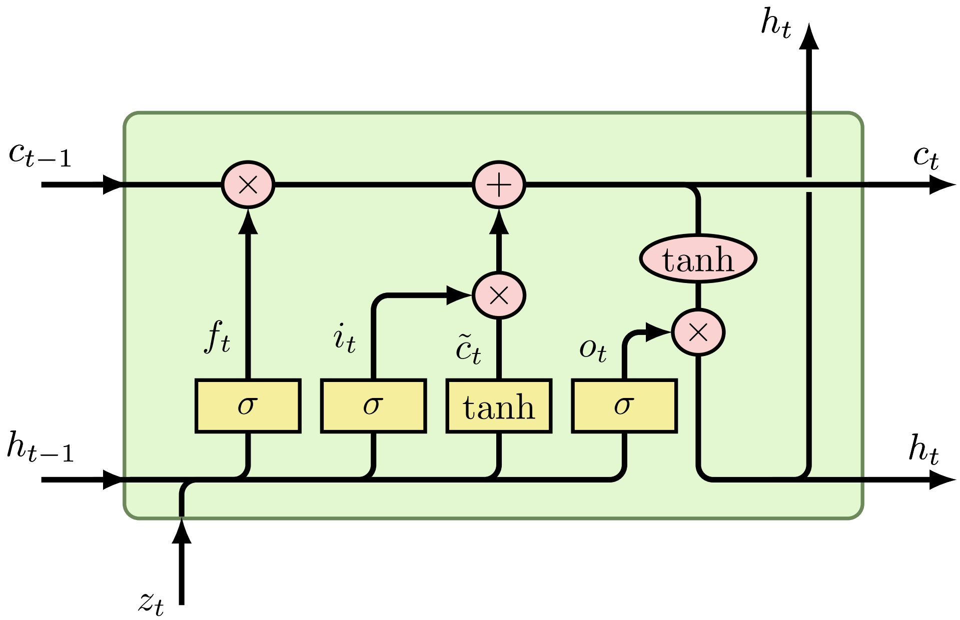
$$\begin{aligned} \text{input}_t &= g\left(W_h^{(i)} h_{t-1} + W_z^{(i)} z_t + w_0^{(i)}\right), \\ \text{forget}_t &= g\left(W_h^{(f)} h_{t-1} + W_z^{(f)} z_t + w_0^{(f)}\right), \\ \text{out}_t &= g\left(W_h^{(o)} h_{t-1} + W_z^{(o)} z_t + w_0^{(o)}\right). \end{aligned} \tag{18}$$

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

$$\begin{aligned} c_t &= \text{forget}_t \odot c_{t-1} + \text{input}_t \odot \tilde{c}_t \\ h_t &= \text{out}_t \odot \tanh(c_t) \end{aligned} \tag{19}$$

Figure 3 presents the diagram of a long short-term memory network. We consider one hidden layer LSTM method for our model comparison.

Figure 3. Long Short-Term Memory Networks



2.7 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, LSTM. We equally combine $\hat{r}_{i,t+1}^{(m)}$ to obtain a new prediction of expected return,

$$\hat{r}_{i,t+1} = \frac{1}{M} \sum_{m=1}^{M} \hat{r}_{i,t+1}^{(m)}. \tag{20}$$

The idea behind is that the $\hat{r}_{i,t+1}^{(m)}$ may have high variances, equally weighting them can reduce the variance dramatically, although it may increase the bias to some extent. The financial literature, such as **?** and **?**, shows that this method works well for return predictability.

2.8 Performance Evaluation

Following **?**, we use the out-of-sample R-square as the performance metric to assess the predictive performance for individual excess bond return forecasts,

$$R_{OS}^2 = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3} \left(r_{i,t+1} - \hat{r}_{i,t+1}\right)^2}{\sum_{(i,t)\in\mathcal{T}_3} r_{i,t+1}^2}. \tag{21}$$

The R_{OS}^2 statistic pools prediction errors across firms 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. To estimate the out-of-sample R_{OS}^2, we follow **?** and the most common 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 performance, which represents the truly out-of-sample evaluation of the performance.

We use the mean squared forecast error (MSFE)-adjusted statistic of **?** to test the statistical significance of R_{OS}^2. Considering the potentially strong cross-sectional dependence among individual excess bond returns, following **?**, 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. **?** show that this test has an asymptotically standard normal distribution when comparing forecasts even from nested models. 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: $R_{OS}^2 \leq 0$ against $H_A : R_{OS}^2 > 0$).

To compare the out-of-sample predictive accuracies between two methods, we follow **?** and use the modified **?** test, which takes into account of 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

$$\text{DM}_{12} = \bar{d}_{12}/\hat{\sigma}_{\bar{d}}, \tag{22}$$

where $\bar{d}_{12}$ and $\hat{\sigma}_{\bar{d}}$ are 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 $t+1$,

$$d_{12,t+1} = \frac{1}{n_{3,t+1}} \sum_{i=1}^{n_3} \left(\left(\hat{e}_{i,t+1}^{(1)}\right)^2 - \left(\hat{e}_{i,t+1}^{(2)}\right)^2 \right), \tag{23}$$

where $\hat{e}_{i,t+1}^{(1)}$ and $\hat{e}_{i,t+1}^{(2)}$ are the return forecast errors of individual asset i at time $t+1$ using each method, 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 a summary 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.[7] We then merge corporate bond pricing data with the Mergent fixed income securities database to obtain bond characteristics such as offering amount, offering date, maturity date, coupon rate, coupon type, interest payment frequency, bond type, bond rating, bond option features, and issuer information.

For TRACE data, we adopt the filtering criteria proposed by ?. Specifically, we re-

[7]We use enhanced TRACE compared to the widely employed 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 tothe 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.

move bonds that (i) are not listed or traded in the U.S. public market; (ii) are structured notes, mortgage-backed, asset-backed, agency-backed, or equity-linked; (iii) are convertible; (iv) trade under \$5 or above \$1,000; (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, 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.

3.1 Corporate Bond Return

The monthly corporate bond return at time t is computed as

$$R_{i,t} = \frac{P_{i,t} + AI_{i,t} + C_{i,t}}{P_{i,t-1} + AI_{i,t-1}} - 1. \tag{24}$$

where $P_{i,t}$ is the transaction price, $AI_{i,t}$ is accrued interest, and $C_{i,t}$ is the coupon payment, if any, of bond i in month t. We denote $r_{i,t}$ as bond i's excess return, $r_{i,t} = R_{i,t} - r_{f,t}$, where $r_{f,t}$ 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 **?**, who discuss the conversion methods in 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.

3.2 Corporate Bond Characteristics and Stock Characteristics

We first build a comprehensive data library of 41 corporate bond characteristics that are motivated by the existing studies on the cross-section of corporate bond returns literature. This broad set of bond characteristics 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, and (v) past bond return characteristics such as bond momentum, short-term reversal, and distributional characteristics including return volatility, skewness, and kurtosis. Appendix of the paper gives a detailed list of the 41 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.

Since we also investigate the predictive power of stock and bond characteristics for future *stock* returns, we obtain monthly individual common equity returns from CRSP for all firms listed in the NYSE, AMEX, and NASDAQ. We also obtain the one-month Treasury bill rate to calculate stock excess returns. We rely on a large set of 94 stock-level predictive characteristics based on **?**.[8] We restrain our stock sample to begin from July 2002 and ends in December 2017 because we focus on the common sample period when our bond return and characteristics construction become available in TRACE which starts in July 2002.

Our final sample includes 22,941 bonds issued by 6,051 unique firms, yielding a total of 1,197,702 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 sample contains bonds with an average rating of 8.36 (i.e., BBB+), an average issue size of $443 million, and an average time-to-maturity of 9.83 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

[8] Details on each of the 94 firm characteristics can be found in the Appendix in **?** and Table A.6 in **?**.

for the bond-level characteristics and several 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 0.38, 0.32, and 0.46. The bond market beta, β^{Bond}, is also positively associated with rating and illiquidity, with respective correlations of 0.08 and 0.09. Bond maturity and duration are positively correlated with all risk measures, implying that bonds with longer maturity or duration (i.e., higher interest rate risk) have higher β^{Bond}, higher VaR, and higher ILLIQ. Bond size is negatively correlated with VaR and ILLIQ, indicating that bonds with smaller size have higher VaR and higher ILLIQ. The correlations between size and rating and between size and maturity are economically weak.

4 Corporate Bond Characteristics, Stock Characteristics, and the Expected Bond Returns

4.1 Out-of-sample Performance Using Bond Characteristics

Following **?**, we compare and evaluate a variety of machine learning methods including OLS with all covariates, principal component analysis (PCA), partial least square (PLS), LASSO, ridge regression (Ridge), elastic net (ENet), random forest (RF), and feedforward neural network (FFN). On top of these methods, we also add a long-term memory neural network (LSTM) to capture a long memory effect (**?**; **?**). Moreover, we rely on the forecast combination method (Combination) which averages individual expected return forecasts from the aforementioned eight machine learning models (**?**; **?**). The comparison of different machine learning techniques is based on their out-of-sample predictive performance.

Table 2 presents the monthly out-of-sample R-squared (R^2_{OS}, in percentage) for the entire pooled sample of corporate bonds using all 41 bond characteristics listed in the Appendix as the covariates. The first row of Table 2 reports R^2_{OS} for the entire sample of corporate bonds. The first row shows that the OLS model with all 41 bond characteristics produces an R^2_{OS} of -6.76%, 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 Table 2 show that machine learning models substantially improve the R^2_{OS}. 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 3.26% and 3.13% per month, 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 2.45%, 2.77%, and 2.90% per month, 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 3.88% per month using random forests (RF). In addition to nonparametric regressions, we also 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 predictive power 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 model produce a significant R^2_{OS} of 4.40% and 3.82% per month, respectively. Finally, the last column of Table 2 shows that the forecast combination model (Combination) significantly improves the R^2_{OS} to 4.05% per month.

In addition to the full sample of bonds, we also examine the predictive performance of different models for investment-grade and non-investment-grade bonds, respectively, in the second and third row of Table 2. We obtain similar patterns in R^2_{OS} for investment-grade and non-investment-grade bonds for all machine learning models. Specifically, linear models such as OLS perform poorly, with R^2_{OS} -5.23% for investment-grade bonds and even worse for non-investment-grade bonds at -12.38% per month, whereas all of the machine learning models deliver significantly positive R^2_{OS}. The predictive power of machine learning models is higher for investment-grade bonds, with R^2_{OS} in the range of 3.15% (LASSO) to 4.81% (FFN), than for non-investment-grade bonds, whose R^2_{OS} range from 0.01% (Ridge) and 2.94% (FFN).

Finally, we plot the monthly R^2_{OS} (y-axis) for all bonds, investment-grade, and non-investment-grade bonds in Panel A of Fig. 4, which clearly highlights the superior performance of machine learning methods to the linear OLS models.

To make pairwise comparisons of methods, we use the ? test for differences in out-of-sample predictive accuracy between two models. Table 3 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 Table 3 shows a positive and statistically significant test statistic for all the machine learning models, in the range of 3.73 to 4.96, compared to the unconstrained OLS model. First, the dimension reduction models such as PCA and PLS, the penalized methods using LASSO, Ridge, and ENet approaches, and the neural network models including RF and LSTM – produce statistically significant improvements over the unconstrained OLS model. Second, 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 Table 3 shows that the forecast combination model (Combination) produces large and significant statistical improvements over most individual machine learning models.

4.2 Which Bond Characteristics Matter?

In this section, we aim to identify corporate bond characteristics that are important for the expected bond returns while simultaneously controlling for the many other predictors. Following the ranking approach in ? and ?, 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 average these into a single importance measure for each predictor. Fig. 5 reports the resulting importances of the top 10 bond-level characteristics for each method, whereas Fig. 6 reports overall rankings of characteristics for all models.[9]

Figures 5 and 6 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) or expected shortfall (ES), total return volatility (VOL), and systematic risk related to bond market beta, default and term beta (β^{Bond}, β^{DEF}, and β^{TERM}), (iii) bond-level illiquidity measures such as average bid and ask price (AvgBidAsk), and Amihud or Roll's measures of illiquidity, (iv) bond return characteristics related to bond momentum (MOM) and short-term reversal (REV). Fig. 5 shows that risk measures play 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 random forest model rely more heavily on bond-level illiquidity measures such as 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 model shows that bond momentum (MOM), return volatility (VOL), and illiquidity (ILLIQ) are the top important covariates for the predictive performance.

4.3 Machine Learning Portfolios Using Bond Characteristics

To further investigate the economic significance of machine learning approaches, we form portfolios based on machine learning forecasts using the 41 bond characteristics. At the end of each month, we calculate one-month-ahead out-of-sample bond return predictions for each of the ten methods. We then sort bonds into deciles based on each model's forecasts and construct the equal-weighted portfolio based on the out-of-sample forecasts.[10] "Low"

[9]The color gradient within each column in Fig. 6 shows the model-specific ranking of characteristics, where the lightest (darkest) presents the least (most) important bond characteristics within each model.

[10]We first report results using the equal-weighted portfolio because our statistical objective functions minimize equally weighted forecast errors. To mitigate the concern of size and illiquidity driven by small bonds, Table A.XX of the online appendix also reports the value-weighted portfolio results using amount outstanding as weights. Our results remain similar.

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.

Table 4 reports the monthly performance results. Panel A shows the results for all sample of bonds. Consistent with our earlier findings using the out-of-sample R-squared as the performance metrics, Panel A of Table 4 shows that all machine learning forecasts generate economically and statistically significant return spreads between High and Low bond portfolios, in the range of 0.58% to 1.20% per month, compared to the unconstrained OLS model which delivers the smallest return spread of 0.47%. The top two best high minus low strategy come from the forecast combination model (Combination) and the feed forward neural networks (FFN), with monthly return spread of 1.20% and 0.98%, respectively.

Next, we examine the performance of machine learning bond portfolios for investment-grade and non-investment-grade bonds, respectively, in Panels B and C of Table 4. Panel B shows that the High minus Low return spreads remain similar for investment-grade bonds, in the range of 0.58% and 1.20% per month. The largest improvement of machine learning portfolios are presented in Panel C for non-investment-grade bonds. The first row of Panel C shows an economically small return spread of 0.11% using the unconstrained OLS forecasts, indicating that linear model such as OLS fail to capture the nonlinear payoffs of corporate bonds which is more pronounced among the non-investment-grade group. However, the return spreads associated with the nine machine learning models are economically large ranging from 0.25% to 1.13%. Overall, Table 4 shows that machine learning approaches significantly improve the forecast performance for bond portfolios using bond characteristics as the covariates.

4.4 Out-of-sample Performance 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. Motivated by this observation, a few studies investigate whether a variety of stock characteristics impact corporate bond returns using cross-sectional Fama-MacBeth regressions (Chordia et al., 2017; Choi and Kim, 2016). These studies find mixed evidence on the role of stock characteristics for the expected bond returns.[11] Compared to these studies which draw from the well of a few predictors, we extend the list of candidate 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.

Table 5 presents the monthly out-of-sample R-squared (R^2_{OS}, in percentage) for the entire pooled sample of corporate bonds using all 94 firm and stock characteristics from **?** and **?** as the covariates. The first row of Table 5 shows that the OLS model with all 94 stock characteristics produces an R^2_{OS} of -2.35%, 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 Table 5 show that machine learning models substantially improve the R^2_{OS}. The penalized methods approach (LASSO, Ridge, and ENet) generate an R^2_{OS} of 2.45%, 2.55%, and 2.46% per month, respectively, higher than 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 0.34% and 1.61%, respectively. Panel B of Fig. 4 plots the monthly R^2_{OS} associated with stock characteristics and shows that the R^2_{OS} is in the range of 0.34% to 2.75%, which is about half of those generated by using corporate bond characteristics in Panel A of Fig. 4.

Next, we examine the predictive performance of different models for investment-grade and non-investment-grade bonds, respectively, using the 94 stock characteristics. The numbers in Panels B and C of Table 5 show similar pattern to those in Panel A for all bonds. Specifically, linear models such as OLS perform poorly, with R^2_{OS} -2.34% for investment-grade bonds and -2.35% for non-investment-grade bonds, whereas all machine learning models deliver significantly positive R^2_{OS}. Interestingly, for each machine learning method, the predictive performance is higher for non-investment-grade bonds in Panel C, with R^2_{OS} in the range of

[11] 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 (2016) 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.

0.36% (FFN) and 3.22% (Combination), than for investment-grade bonds in Panel B, whose R^2_{OS} range from 0.22% (FFN) and 2.61% (Combination).

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

Our previous sections investigate the performance of machine learning approaches using the 41 bond characteristics and 94 stock characteristics separately. The results so far suggest that all machine learning models produce significantly positive predictive power using either set of characteristics, although the predictive performance associated with using bond characteristics is twice as strong as using stock characteristics. In this section, we answer the question on whether stock characteristics provide incremental predictive power for the expected bond returns relative to bond characteristics.

Table 6 reports R^2_{OS} combining the 41 bond characteristics and 94 stock characteristics (i.e., 135 characteristics in total). Consistent with our previous findings, the unconstrained OLS model produces R^2_{OS} of -3.72%, indicating that the model fails to deliver statistically significant out-of-sample forecasting power for the expected corporate bond returns. The other columns of Table 6 show that machine learning models using the combined 135 characteristics deliver significantly positive R^2_{OS} ranging from 1.69% to 3.67%. The results remain similar when we investigate investment-grade and non-investment-grade bonds, respectively.

To evaluate the marginal improvement in predictive performance, Panel B of Table 6 reports the Diebold-Mariano test statistics for pairwise comparisons of a column model versus a row model, where the column model uses the combined 135 stock plus bond characteristics and the row model uses only the 41 bond characteristics. A significantly positive statistic indicates that stock characteristics improve the predictive power of bond characteristics in forecasting expected bond returns. However, Panel B of Table 6 shows that none of the test statistics is significantly positive, indicating that there is no difference in the performance of machine learning models when adding stock characteristics to the bond characteristics in forecasting bond returns. Finally, Panel C of Fig. 4 plots the monthly R^2_{OS} using both stock and bond characteristics and shows similar results to those in Panel A using bond characteristics only. Overall, we conclude that although stock characteristics produce significant explanatory power for bond returns when used alone, their incremental improvement relative to bond characteristics is small.

5 Do Stock or Bond Characteristics Predict Future Stock Returns?

? find that machine learning offers an improved description of expected return relative to traditional methods in forecasting future stock returns. Motivated by ?, in this section, we examine the predictive performance of the 94 stock and 41 bond characteristics, respectively, in predicting future *stock* returns. Consistent with the findings in ?, we show that machine learning methods provide strong forecasting power using the stock characteristics. Moreover, we uncover new evidence that when used alone in the machine learning models, corporate bond characteristics perform well in predicting future stock returns. However, when comparing the marginal improvement in predictive performance, we find that bond characteristics do not offer additional explanatory power above and beyond stock characteristics.

5.1 Out-of-Sample Performance

Table 7 presents the monthly out-of-sample R^2_{OS} for the entire pooled sample of stocks using all 94 stock characteristics and 41 bond characteristics, respectively, as the covariates. Following ?, we obtain monthly common equity returns from CRSP for all firms listed in the NYSE, AMEX, and NASDAQ. We restrain our sample to begin from July 2002 and ends in December 2017 because we focus on the common periods when our bond sample construction and bond characteristics become available. The first row of Table 7 shows that the OLS model with all 94 stock characteristics produces an R^2_{OS} of -3.83%, indicating that the model fails to deliver statistically significant out-of-sample forecasting power for the expected stock returns.[12] Consistent with the findings in ?, machine learning methods significantly improve

[12]This result is consistent with ?, who finds an out-of-sample R-squared of -3.46 for the OLS model over an extended sample period using similar stock characteristics from March 1957 to December 2016.

the out-of-sample forecasting power for the expected stock returns, with monthly R^2_{OS} ranging from 0.08% to 0.60%. Specifically, forecast combination method (Combination) and neutral network models (FFN and LSTM) deliver the highest R^2_{OS} of 0.60%, 0.56%, and 0.54%, respectively. Regression trees (RF) and dimension reduction models (PCA and PLS) deliver the second largest improvement of the R^2_{OS} to 0.49%, 0.35% and 0.44%, respectively, relative to the OLS model. The penalized methods (LASSO, Ridge, and ENet) generate the third largest improvement over the OLS model, with R^2_{OS} in the range of 0.08% and 0.12%.

The second row of Table 7 uses all 41 bond characteristics as the covariates, and shows that machine learning methods using corporate bond characteristics are able to generate significantly positive R^2_{OS} ranging from 0.11% to 0.44% per month. Fig. 7 plots the monthly R^2_{OS} (y-axis) for all models using bond and stock characteristics, respectively, and shows that corporate bond characteristics deliver similar R^2_{OS} to those of stock characteristics.

To what extent do bond characteristics provide incremental predictive power for future stock returns, compared to using stock characteristics alone? To answer this question, we investigate the joint predictive power of the 94 stock and 41 bond characteristics in forecasting future stock returns in Table 8. Panel A of Table 8 shows that machine learning models using the combined 135 characteristics deliver significantly positive R^2_{OS} ranging from 0.23% to 0.65%. However, as shown in Panel B, most of the Diebold-Mariano test statistics are insignificant for the machine learning methods, which indicates that the column model (i.e., the combined 135 stock and bond characteristics) does not significantly outperform the row model that uses only the 94 stock characteristics. Overall, Table 8 suggests that there is no significant improvement in the predictive performance when adding bond characteristics to the stock characteristics in forecasting stock returns.

6 Conclusion

Using a variety of machine learning methods, we provide a comprehensive study of the expected corporate bond returns using a large set of 94 stock characteristics and 41 bond characteristics. Both equity and corporate bonds are contingent claims on firms but they also differ in several key features. Thus it is important to examine whether stock characteristics or bond characteristics, individually or jointly, drive the expected corporate bond returns. 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 relationship.

We find that traditional unconstrained linear models such as OLS perform poorly, whereas machine learning methods substantially improve the out-of-sample forecasting power for the expected corporate bond returns. In addition, the predictive power of machine learning models extend to both investment-grade and non-investment-grade bonds. While stock characteristics produce significant explanatory power for bond returns when used alone, their predictive power is only half of that associated with using corporate bond characteristics, and their incremental improvement relative to bond characteristics is small.

Finally, in addition to bond return predictability, we investigate whether the comprehensive list of stock and bond characteristics, individually or jointly, predict future stock returns. We find that when used alone as the predictive covariates, bond characteristics provide as strong forecasting power for future stock returns as using stock characteristics alone. However, bond characteristics do not offer additional explanatory above and beyond stock characteristics when we combine both set of characteristics.

Appendix: Corporate Bond Characteristics

The Appendix describes the comprehensive data library of the 41 corporate bond characteristics we construct, which are motivated by the existing studies on the cross-section of corporate bond returns literature. This broad set of bond characteristics can be largely classified to (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, and (v) past bond return characteristics such as bond momentum, reversal, and distributional characteristics such as return volatility, skewness, and kurtosis. The 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.

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*).**

6. **Downside risk proxied by the 5% VaR (*VaR5*).** Following **?**, 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 (VaR), 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.[13]

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 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

[13]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.

illiquidity using daily bond returns, the Roll (1984) measure is defined as,

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

where r_d is the corporate bond return on day d.

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

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

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

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

$H_t(L_t)$ 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,

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

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

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

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,

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

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

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

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

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

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*).**

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

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 et al. (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:

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

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:

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

We calculate 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 Feldhutter (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})**. Pastor 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:

$$r_{t+1}^e = \theta + \psi \cdot r_t + \gamma \cdot sign(r_t^e) \cdot Q_t + \epsilon_t, \tag{25}$$

where r_t^e is the security'e 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.[14]

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).

30. **Downside risk beta (β^{DRF}).** Following **?**, 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}).** Following **?**, for each bond and each month in our sample, we estimate the factor beta from the monthly rolling regressions of excess bond returns

[14] 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.

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}).** Following **?**, 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 (2018), 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. **Short-term reversal (*REV*).** The bond return in previous month.

35. **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.

36. **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.

37. **Volatility (*VOL*).** Following Bai, Bali, and Wen (2016), it is estimated using a 36-month rolling window for each bond in our sample

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

38. **Skewness (*SKEW*)**. Following Bai, Bali, and Wen (2016), it is estimated using a 36-month rolling window for each bond in our sample

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

39. **Kurtosis (*KURT*)**. Following Bai, Bali, and Wen (2016), it is estimated using a 36-month rolling window for each bond in our sample

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

40. **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.

41. **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.

References

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

Bai, J., Bali, T., Wen, Q., 2016. Do the distributional characteristics of corporate bonds predict their future returns? Working Paper, SSRN E-Library.

Bao, J., Pan, J., Wang, J., 2011. The illiquidity of corporate bonds. Journal of Finance 66, 911–946.

Bessembinder, H., Kahle, K. M., Maxwell, W. F., Xu, D., 2009. Measuring abnormal bond performance. Review of Financial Studies 22, 4219–4258.

Bessembinder, H., Maxwell, W. F., Venkataraman, K., 2006. Market transparency, liquidity externalities, and institutional trading costs in corporate bonds. Journal of Financial Economics 82, 251–288.

Boyer, B., Mitton, T., Vorkink, K., 2010. Expected idiosyncratic skewness. Review of Financial Studies 23, 169–202.

Chakravarty, S., Sarkar, A., 2003. Trading costs in three u.s. bond markets. Journal of Fixed Income 13, 39–48.

Choi, J., Kim, Y., 2016. Anomalies and market (dis)integration. Working Paper, SSRN eLibrary .

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

Chung, K. H., Wang, J., Wu, C., 2018. Volatility and the cross-section of corporate bond returns. Journal of Financial Economics, forthcoming.

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

Dick-Nielsen, J., Feldhutter, P., Lando, D., 2012. Corporate bond liquidity before and after theonset of the subprime crisis. Journal of Financial Economics 103, 471–492.

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

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

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

Fong, K., Holden, C. W., Trzcinka, C. A., 2017. What are the best liquidity proxies for global research? Review of Finance 21, 1355–1401.

Gebhardt, W. R., Hvidkjaer, S., Swaminathan, B., 2005. The cross section of expected corporate bond returns: betas or characteristics? Journal of Financial Economics 75, 85–114.

Goyenko, R., Holden, C., Trzcinka, C., 2009. Do liquidity measures measure liquidity? Journal of Financial Economics 92, 153–181.

Graham, J., Leary, M., Roberts, M., 2015. A century of capital structure: The leveraging of corporate america. Journal of Financial Economics 118, 658–683.

Han, S., Zhou, H., 2007. Nondefault bond spread and market trading liquidity. Working Paper, Federal Reserve Board.

Harvey, C. R., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55, 1263–1295.

Hasbrouck, J., 2009. Trading costs and returns for u.s. equities: Estimating effective costs from daily data. Journal of Finance 65, 1445–1477.

Hong, G., Warga, A., 2000. An empirical study of bond market transactions. Financial Analyst Journal 56, 32–46.

Hong, H., Sraer, D., 2013. Quiet bubbles. Journal of Financial Economics 110, 596–606.

Jostova, G., Nikolova, S., Philipov, A., Stahel, C., 2013. Momentum in corporate bond returns. Review of Financial Studies 26, 1649–1693.

Lesmond, D., Ogden, J., Trzcinka, C., 1999. A new estimate of transaction costs. Review of Financial Studies 12, 1113–1141.

Lin, H., Wang, J., Wu, C., 2011. Liquidity risk and the cross-section of expected corporate bond returns. Journal of Financial Economics 99, 628–650.

Lo, A. W., MacKinlay, A. C., 1990. When are contrarian profits due to stock market overreaction? Review of Financial Studies 3, 175–205.

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

Pastor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

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

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

Schestag, R., Schuster, P., Uhrig-Homburg, M., 2016. Measuring liquidity in bond markets. Review of Financial Studies 29, 1170–1219.

Table 1 Descriptive statistics

Panel A reports the number of bond-month 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}. 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	1st	5th	25th	75th	95th	99th
Bond return (%)	1,197,702	0.79	0.50	5.90	-9.17	-4.19	-0.73	1.91	6.37	14.06
Rating	1,197,702	8.36	7.77	3.95	1.60	2.38	5.62	10.38	16.12	18.97
Time to maturity (maturity, year)	1,197,702	9.83	6.68	9.20	1.16	1.60	3.70	13.36	26.92	32.36
Amount out (size, $million)	1,197,702	442.53	300.24	532.44	4.21	17.13	109.03	562.14	1391.03	2570.22
Duration (DUR)	810,064	6.00	5.00	3.82	0.99	1.45	3.11	7.82	13.62	14.98
Downside risk (5% VaR)	629,375	5.66	3.96	5.63	0.69	1.14	2.39	6.74	16.20	28.85
Illiquidity (IILIQ)	986,377	2.04	0.43	5.03	-1.11	-0.21	0.07	1.88	9.56	23.01
Bond market beta (β^{Bond})	634,809	1.02	0.84	0.87	-0.45	0.07	0.48	1.34	2.67	3.89

Panel B: Average cross-sectional correlations

	Rating	Maturity	Size	DUR	VaR	ILLIQ	β^{Bond}
Rating	1	-0.13	-0.04	-0.19	0.46	0.13	0.08
Maturity		1	-0.04	0.89	0.17	0.11	0.35
Size			1	-0.02	-0.10	-0.15	0.07
DUR				1	0.25	0.11	0.49
VaR5					1	0.32	0.38
ILLIQ						1	0.09
β^{Bond}							1

Table 2 Do corporate bond characteristics predict corporate bond returns?

This table reports monthly out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of corporate bonds using the 41 bond characteristics. 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,

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

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%.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Rdige	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
All bonds	−6.76	3.26	3.13	2.45	2.77	2.90	3.88	4.40	3.82	4.05
Investment-grade bonds	−5.23	3.84	3.74	3.15	3.54	3.62	4.23	4.81	4.12	4.63
Non-investment-grade bonds	−12.38	1.43	1.25	0.03	0.01	0.04	2.43	2.94	2.81	2.10

Table 3 Comparison of monthly out-of-sample prediction using Diebold-Mariano tests

This table reports pairwise Diebold-Mariano test statistics comparing the out-of-sample bond-level prediction performance (R^2_{OS}) among the ten 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.

	OLS	PCA	PLS	LASSO	Ridge	ENet	RF	FFN	LSTM	Combination
OLS	–	**3.97**	**3.73**	**4.70**	**4.96**	**4.46**	**4.27**	**4.94**	**3.94**	**4.76**
PCA		–	−0.59	−0.75	−0.36	−0.40	0.63	−0.33	−0.08	0.96
PLS			–	−0.51	−0.17	−0.18	0.70	−0.18	0.06	1.01
LASSO				–	**1.98**	1.00	1.44	0.52	0.66	**3.26**
Ridge					–	0.00	1.02	−0.15	0.26	**2.21**
ENet						–	1.30	−0.14	0.36	**4.27**
RF							–	−0.99	−0.98	−0.04
FFN								–	0.36	**1.56**
LSTM									–	**1.09**
Combination										–

Table 4 Performance of machine learning bond portfolios using corporate bond characteristics

This table reports the monthly performance of equal-weighted decile portfolios sorted on out-of-sample machine learning return forecasts using the 41 bond characteristics (i.e., $\hat{r}_{i,t+1}$ where $(i,t) \in \mathcal{T}_3$, the test subsample). At the end of each month, we calculate one-month-ahead out-of-sample bond return predictions for each method. We then sort bonds into deciles based on each model's forecasts and construct the equal-weighted portfolio 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.

Panel A: All bonds

	Low	2	3	4	5	6	7	8	9	High	High−Low	t-stat
OLS	1.07	0.80	0.86	0.98	0.98	1.16	1.11	0.81	1.35	1.54	0.47	2.34
PCA	0.64	0.61	0.59	0.62	0.76	0.82	0.85	0.88	0.93	1.32	0.68	3.97
PLS	0.75	0.59	0.61	0.59	0.69	0.78	0.81	0.86	0.92	1.33	0.58	4.23
LASSO	0.78	0.76	0.67	0.58	0.63	0.62	0.67	0.74	0.91	1.37	0.59	2.75
Ridge	0.39	0.42	0.41	0.38	0.43	0.52	0.55	0.60	0.75	1.11	0.72	3.05
ENet	0.57	0.46	0.44	0.39	0.43	0.48	0.51	0.66	0.74	1.06	0.49	4.03
RF	0.62	0.72	0.72	0.60	0.68	0.70	0.68	0.73	0.91	1.48	0.85	3.23
FFN	0.45	0.53	0.56	0.53	0.65	0.75	0.85	0.94	1.11	1.42	0.98	3.64
LSTM	0.62	0.52	0.50	0.52	0.56	0.62	0.67	0.79	0.90	1.21	0.58	2.78
Combination	0.27	0.41	0.50	0.55	0.62	0.66	0.74	0.84	1.00	1.47	1.20	4.77

Panel B: Investment-grade bonds

	Low	2	3	4	5	6	7	8	9	High	High−Low	t-stat
OLS	1.04	0.78	0.86	0.98	1.07	1.12	1.10	0.93	1.33	1.54	0.50	2.42
PCA	0.65	0.63	0.59	0.62	0.75	0.80	0.85	0.90	0.95	1.33	0.68	3.92
PLS	0.76	0.62	0.60	0.59	0.68	0.77	0.82	0.88	0.92	1.35	0.59	4.30
LASSO	0.78	0.81	0.69	0.58	0.61	0.60	0.68	0.74	0.92	1.40	0.63	3.90
Ridge	0.40	0.44	0.41	0.38	0.42	0.50	0.55	0.61	0.77	1.14	0.74	3.01
ENet	0.59	0.44	0.45	0.38	0.43	0.46	0.50	0.68	0.75	1.08	0.49	4.14
RF	0.62	0.78	0.72	0.59	0.65	0.69	0.67	0.73	0.92	1.51	0.89	3.28
FFN	0.42	0.56	0.57	0.52	0.63	0.74	0.85	0.95	1.15	1.45	1.03	3.21
LSTM	0.62	0.52	0.50	0.52	0.56	0.62	0.67	0.79	0.90	1.21	0.58	4.00
Combination	0.28	0.42	0.51	0.54	0.61	0.67	0.72	0.88	0.98	1.48	1.20	4.60

Table 4 (Continued)

Panel C: Non-investment-grade bonds

	Low	2	3	4	5	6	7	8	9	High	High−Low	t-stat
OLS	1.40	0.89	0.89	0.85	0.76	1.50	0.96	0.64	1.18	1.51	0.11	2.46
PCA	0.68	0.47	0.62	0.53	0.71	0.84	0.76	0.77	0.75	1.31	0.63	2.80
PLS	0.83	0.53	0.64	0.61	0.54	0.71	0.74	0.72	0.76	1.27	0.45	3.00
LASSO	0.68	0.61	0.60	0.56	0.50	0.63	0.54	0.68	0.84	1.44	0.76	2.27
Ridge	0.46	0.28	0.39	0.32	0.20	0.51	0.37	0.51	0.59	1.04	0.59	2.63
ENet	0.63	0.33	0.41	0.30	0.28	0.39	0.36	0.63	0.56	0.89	0.25	2.77
RF	0.63	0.71	0.58	0.59	0.47	0.61	0.64	0.77	0.86	1.50	0.87	2.71
FFN	0.50	0.50	0.51	0.44	0.54	0.65	0.79	0.99	1.01	1.28	0.78	2.78
LSTM	0.76	0.59	0.55	0.52	0.55	0.61	0.72	0.82	0.88	1.21	0.46	2.38
Combination	0.36	0.48	0.56	0.58	0.62	0.61	0.77	0.75	1.02	1.49	1.13	3.23

Table 5 Do stock characteristics predict corporate bond returns?

This table reports monthly out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of corporate bonds 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), 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,

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

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%.

	(1) OLS	(2) PCA	(3) PLS	(4) LASSO	(5) Ridge	(6) ENet	(7) RF	(8) FFN	(9) LSTM	(10) Combination
All bonds	−2.35	2.28	1.78	2.45	2.55	2.46	0.86	0.34	1.61	2.75
Investment-grade bonds	−2.34	2.21	1.54	2.33	2.43	2.35	0.60	0.22	1.41	2.61
Non-investment-grade bonds	−2.35	2.58	2.56	2.83	2.94	2.83	1.81	0.36	2.23	3.22

Table 6 Do stock characteristics improve the predictive power of bond characteristics for bond returns?

Panel A of this table reports monthly out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of corporate bonds combining all bond characteristics and 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,

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

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%. Panel B reports the Diebold-Mariano test statistics comparing the differences in R^2_{OS} between using bond characteristics only and using both bond and stock characteristics.

Panel A: Out-of-sample R-squared (R^2_{OS}) combining both bond and stock characteristics

	OLS	PCA	PLS	LASSO	Ridge	ENet	RF	FFN	LSTM	Combination
All bonds	−3.72	3.14	3.28	2.91	3.15	2.94	2.68	1.69	2.74	3.67
Investment-grade bonds	−3.71	3.48	3.60	3.16	3.40	3.18	2.92	2.00	2.88	3.93
Non-investment-grade bonds	−3.73	2.23	2.37	2.33	2.60	2.38	2.12	0.93	2.50	3.05

Panel B: Comparing R^2_{OS} using Diebold-Mariano test statistics

	Combining both bond and stock characteristics								
	PCA	PLS	LASSO	RIDGE	ENET	RF	FFN	LSTM	Combination
Bond characteristics only	−0.61	0.16	0.66	−0.03	−0.97	−1.05	−0.18	−0.23	−0.43
p-value	0.27	0.44	0.26	0.49	0.17	0.15	0.43	0.41	0.33

Table 7 Do stock or corporate bond characteristics predict future stock returns?

This table reports monthly out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of stocks using the 94 stock characteristics and 41 bond 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,

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

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%.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	PCA	PLS	LASSO	Ridge	ENet	RF	FFN	LSTM	Combination
Stock characteristics only	−3.83	0.35	0.44	0.12	0.08	0.12	0.49	0.54	0.56	0.60
Bond Characteristics only	−3.59	0.14	0.29	0.13	0.11	0.14	0.34	0.30	0.36	0.44

Table 8 Do bond characteristics improve the predictive power of stock characteristics for future stock returns?

Panel A of this table reports monthly out-of-sample R-squared (R^2_{OS}, in percentage) for the entire panel of stocks combining all 94 stock characteristics and 41 bond 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,

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

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%. Panel B reports the Diebold-Mariano test statistics comparing the differences in R^2_{OS} between using stock characteristics only and using both stock and bond characteristics.

Panel A: Out-of-sample R-squared (R^2_{OS}) combining both stock and bond characteristics

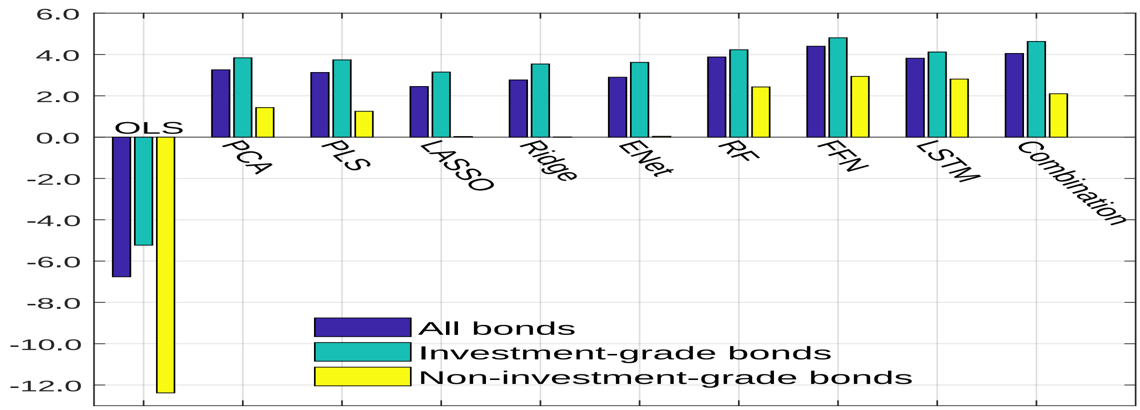
	OLS	PCA	PLS	LASSO	RIDGE	ENET	RF	FFN	LSTM	Combination
Stock and bond characteristics combined	−4.29	0.32	0.42	0.23	0.30	0.24	0.53	0.57	0.56	0.65

Panel B: Comparing R^2_{OS} using Diebold-Mariano test statistics

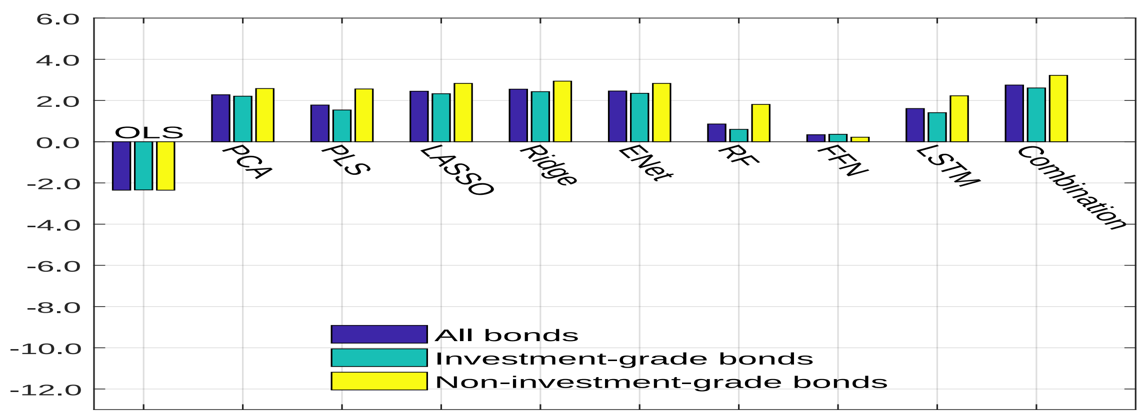
	Combining both stock and bond characteristics								
	PCA	PLS	LASSO	Ridge	ENet	RF	FFN	LSTM	Combination
Stock characteristics only	−0.37	−0.27	0.77	2.50	0.79	1.39	0.28	1.32	0.63
p-value	0.36	0.39	0.22	0.01	0.22	0.08	0.39	0.35	0.27

Figure 4. Monthly Out-of-Sample R^2_{OS} of Corporate Bond Returns

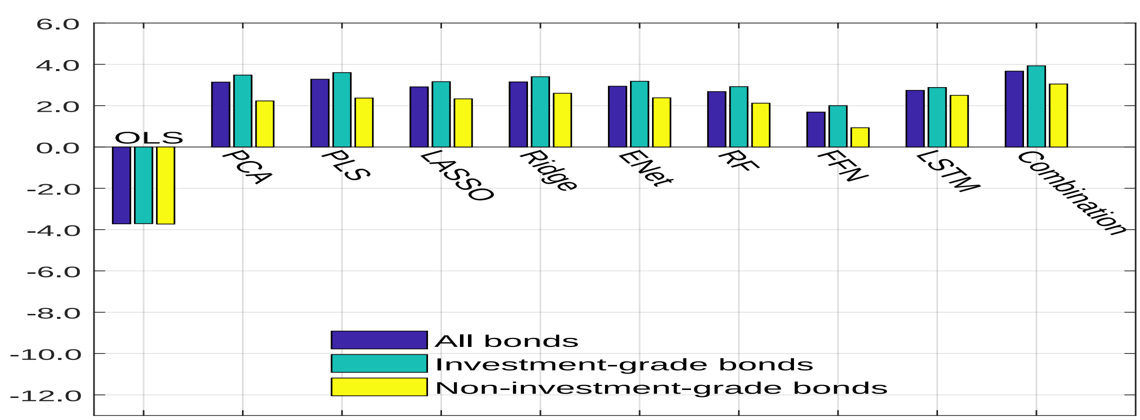
Panel A: Monthly R^2_{OS} Using Bond Characteristics Only



Panel B: Monthly R^2_{OS} Using Stock Characteristics Only

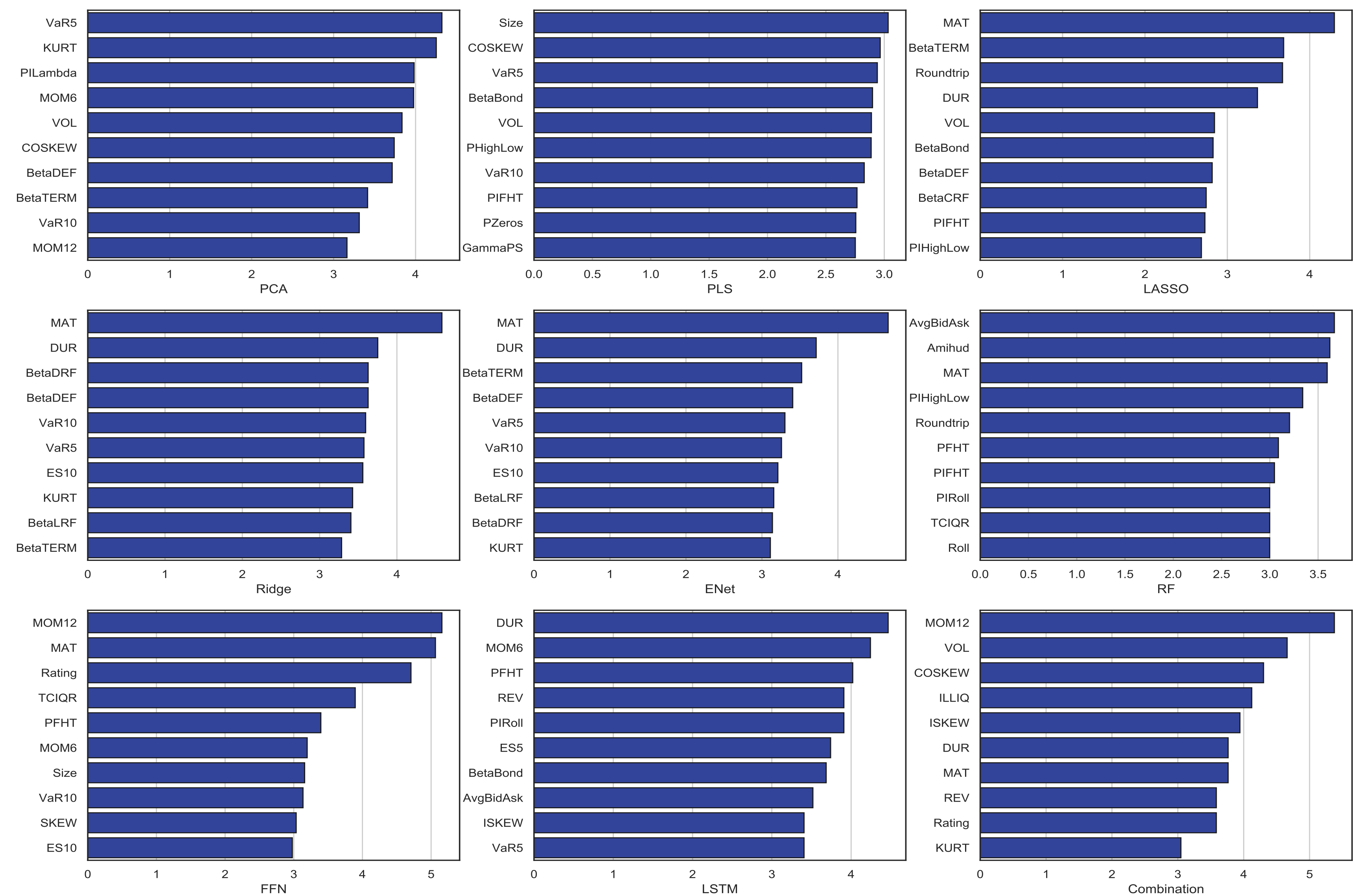


Panel C: Monthly R^2_{OS} Using Both Bond and Stock Characteristics



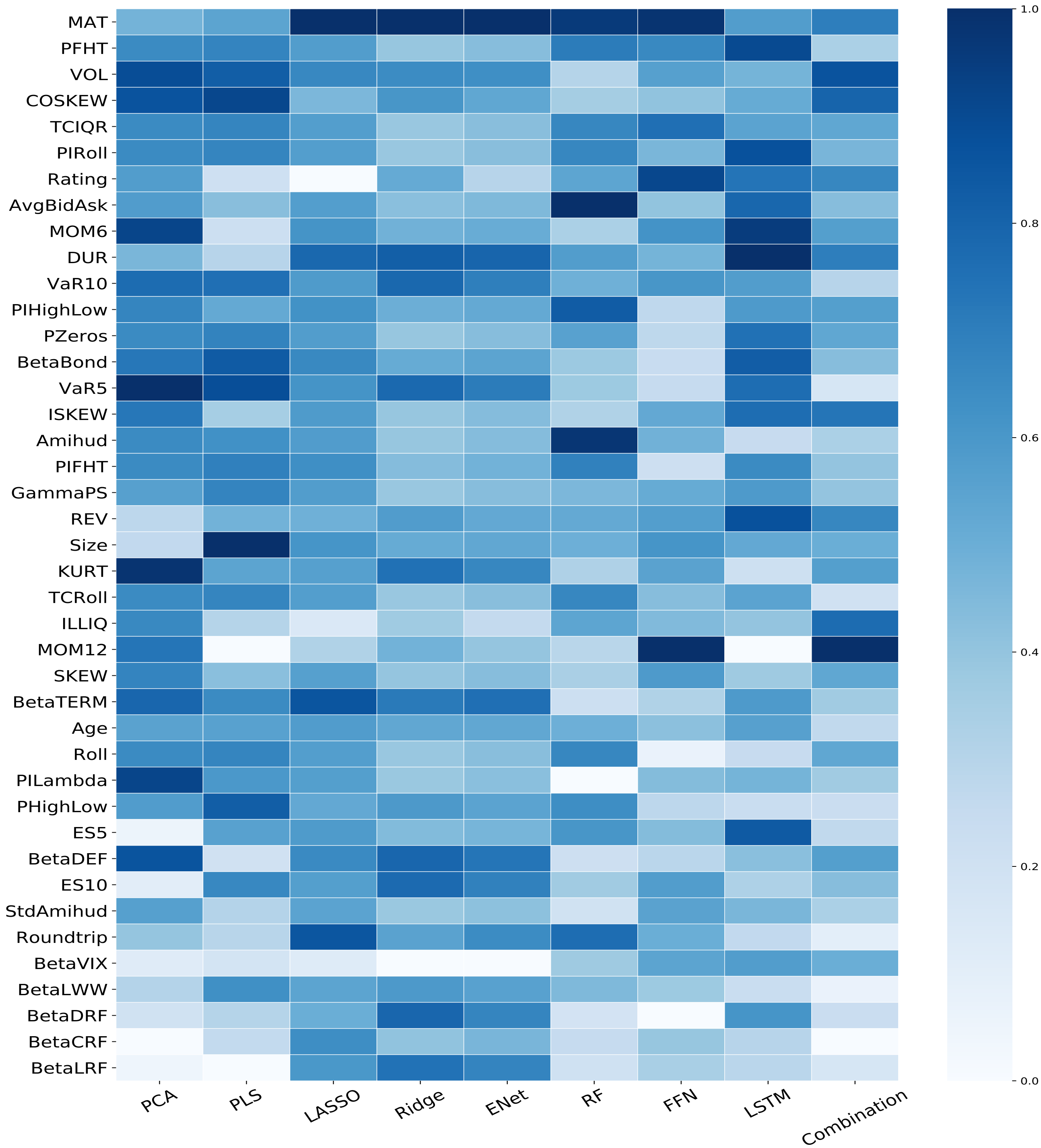
This figure presents the monthly out-of-sample R-squared (R^2_{OS}, in percentage) of corporate bond returns using OLS, principal component analysis (PCA), partial least square (PLS), LASSO, ridge regression (Ridge), elastic net (Enet), random forest (RF), feedforward neural network (FFN), long short-term memory neural network (LSTM), and forecast combination (Combination). Panel A reports R^2_{OS} using only 41 bond characteristics and Panel B reports R^2_{OS} using only 94 stock characteristics. Figure reports the monthly R^2_{OS} for all bonds, investment-grade, and non-investment-grade bonds.

Figure 5. Variable importance by model for corporate bond return prediction



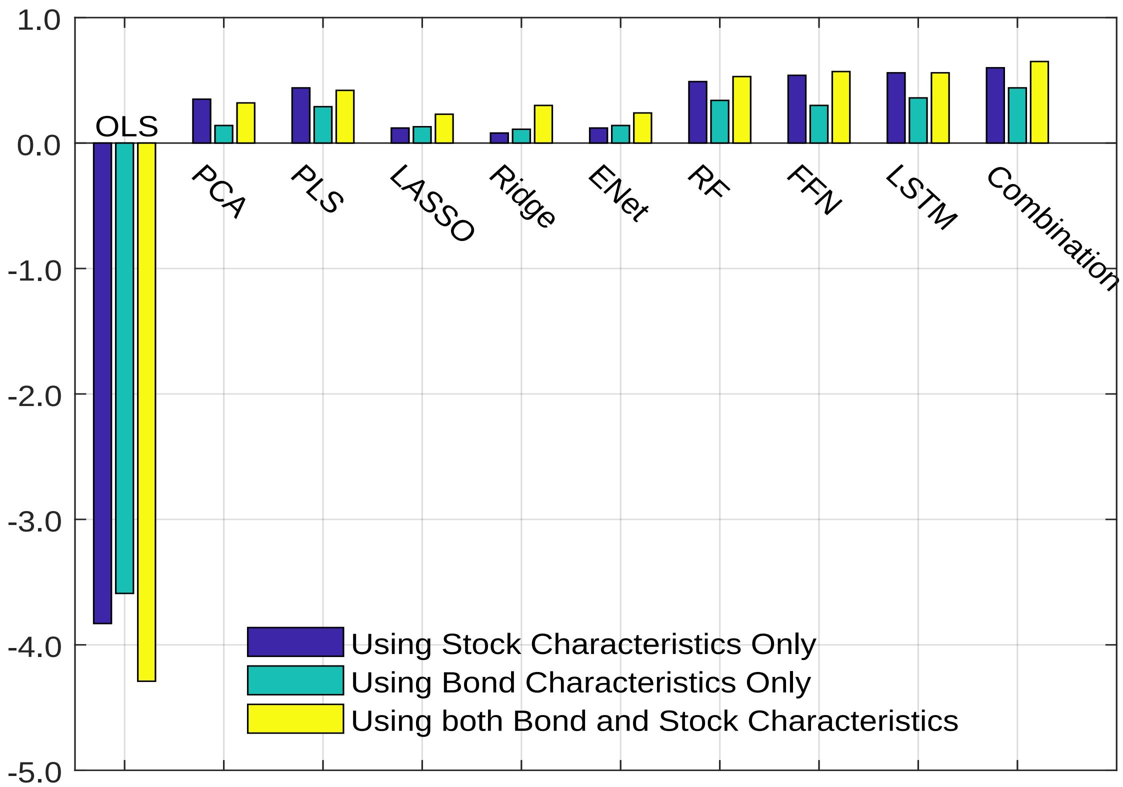
This figure presents the variable importance for the top 10 most influential bond-level characteristics in each model for corporate bond returns, using the 41 bond characteristics as the covariates. For each model, we calculate the reduction in R^2 from setting all values of a given predictor to zero 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 6. Characteristic importance for corporate bond return prediction



This figure presents the overall rankings of bond-level 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 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 7. Monthly Out-of-Sample R^2_{OS} of Stock Returns



This figure presents the monthly out-of-sample R-squared (R^2_{OS}, in percentage) of stock 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 monthly R^2_{OS} using 94 stock characteristics and 41 bond characteristics, respectively.