Predicting Individual Corporate Bond Returns *

Xin He † Guanhao Feng ‡ Junbo Wang § Chunchi Wu ¶

January 22, 2021

Abstract

This paper finds substantial predictability evidence and investment gains for individual corporate bond returns via machine learning. Our predictor set consists of 20 macro predictors and 20 corporate bond characteristics. First, we find significantly positive out-of-sample prediction evidence for individual bond returns in the overall sample and various sub-samples by adopting the Fama-Macbeth approach for machine learning forecasts. Second, a long-short strategy constructed by random forest forecasts delivers an average monthly return of 1.48% and a significant alpha of 1.40% over a five-factor model. Both linear and nonlinear machine learning methods, such as lasso and random forest, are shown to be robust with common predictors, including lag bond market returns, three-month treasury bill rate, and short-term reversal. The lagged bond market return is probably the most important predictor and acts as a short-term reversal for the bond market. Third, the indifferent prediction performance between public and private bonds suggests equity information might be conditionally redundant. We find the inclusion of equity characteristics is possible to deteriorate the predictability evidence. Finally, the predictability source is mainly from the discount rate component instead of the cash flow component of the corporate bond returns.

Key Words: Return Predictability, Big Data and Machine Learning, Individual Corporate Bonds, Bond Characteristics, Macro Predictors, Equity Characteristics, Return Decomposition.

---

*We appreciate insightful comments from Junye Li.

†City University of Hong Kong. E-mail address: xin.he@my.cityu.edu.hk.
‡City University of Hong Kong. E-mail address: gavin.feng@cityu.edu.hk.
§City University of Hong Kong. E-mail address: jwang2@cityu.edu.hk.
¶State University of New York at Buffalo. E-mail address: chunchiw@buffalo.edu.

1 Introduction

The academic literature on predicting corporate bond returns is relatively short, although extensive research in predicting stock returns has been conducted for the last century. One possible explanation is the large proportion of private companies' corporate bonds, whose information is not as publicly accessible as those publicly listed companies for equities. Though investment in corporate bonds plays a more critical role in current institutional investors' portfolios, the best-known pricing metric is still the bond credit rating. Many modeling techniques and methods well developed for equities are waiting to be tested for the bond market.

A current research line studies the prediction of asset returns via machine learning methods; see Gu et al. (2020) for equities and Bianchi et al. (2020) for treasury bonds. Machine learning methods are well known for having high predictive accuracy due to their flexibility. These methods also deal with a high-dimension of predictors and incorporate possible nonlinear and interactive predictor structures. Linear methods, such as lasso and principal component analysis (PCA), have been widely used in the asset pricing literature. Nonlinear methods, such as random forest and deep neural networks, attract considerable attention from academic and industry researchers. In this paper, we follow Feng et al. (2020) and adopt the Fama-Macbeth approach for machine learning forecasts to evaluate the predictability significance to bond return.

The individual corporate bond returns data is highly imbalanced because most bonds only have a few years of observations. Researchers also have proposed a long list of corporate bond characteristics and macro predictors for return predictability and asset pricing performance. However, the predictors' existence and strength can be vastly different for bonds in different credit ratings, as can the duration of their interactive effects. Lin et al. (2014) find positive return predictability evidence for corporate bond portfolios using naive combination forecasts on simple linear regressions. Our study follows the literature and investigates the predictability of individual bond returns, including private company bonds.

This paper has three goals. We study the return predictability of individual corporate bonds via commonly used machine learning methods in the overall sample and various sub-samples. Second, we investigate the important common predictors that drive the return predictability and evaluate their robustness in different linear and nonlinear methods. Third, evaluating the investment performance through bond return predictability is important, and we construct long-short portfolios to assess the risk-adjusted performances.

The empirical study consists of 589,528 bond-month observations covering 19,782 unique bonds for public and private companies from 1976 to 2017. We find significantly positive out-of-sample prediction evidence for individual corporate bond returns in the overall sample. Second, we further show the robustness of predictability in different sub-samples of corporate bonds sorted by credit rating or duration. The overall predictive evidence is strong during the financial crisis, and notably, investment-grade bonds have higher predictability over non-investment-grade bonds in all sub-periods. Finally, the indifferent prediction performance between public and private bonds suggests equity information might be conditionally redundant.

Such robust bond predictability can translate into excellent investment performance. Long-short strategies constructed by lasso and random forest forecasts deliver average monthly returns of 0.96% and 1.48%, annualized Sharpe Ratio 1.94 and 3.16, respectively, and highly significant alphas over a five-factor model (MKT, SMB, HML, TERM, and DEF) used in Fama and French (1993). Both linear and nonlinear machine learning methods, especially for lasso and random forest, are demonstrated with robust return predictability and positive investment performance. For 20 bond characteristics and 20 macro predictors used in the study, we find the lagged bond market returns, three-month treasury bill rate, and short-term reversal are significantly robust in different linear and nonlinear methods. The lagged bond market return is probably the most important predictor for corporate bond returns, which acts as a short-term reversal for the bond market.

The remainder of the paper is organized as follows. Section 1.1 lists the relevant literature and the position of our work. We discuss machine learning applications in section 2. We introduce the data in section 3 and the empirical design in section 4. The empirical findings are provided in section 5. Finally, section 6 concludes and discusses future research directions.

1.1 Related Literature

Our empirical findings contribute to the literature in several important ways. The main contribution is the area of corporate bond return prediction. Lin et al. (2014) document the forward rate, liquidity factor, and credit spread significantly contribute to the predictive power of corporate bond returns, primarily through combination forecasts. Chordia et al. (2017) also find predictability facts with equity characteristics, such as profitability and asset growth, whereas Bai et al. (2019b) confirm the predictability of return volatility and skewness. Lin et al. (2018) find predictability facts for corporate bond portfolio returns with an iterated combination approach.

Our paper is also related to recent working papers on predicting corporate bond returns, including Bali et al. (2020) and Bredendiek et al. (2019). Both papers predict public individual corporate bond returns or create optimal portfolios using both bond and equity characteristics. In contrast, ours focuses on both public and private companies for an extended period. Another recent work of Guo et al. (2020) finds machine learning predictability of corporate bond returns, using yield predictors that capture the yield curve information.

Our paper is also linked to work that analyzes the pricing factors in the cross-sectional expected corporate bond returns. A large body of literature studies corporate bond risk factors. Lin et al. (2011) find liquidity risk is an important determinant, and Chung et al. (2019) document the volatility risk. Bai et al. (2019a) introduce common risk factors based on corporate bonds' general risk characteristics, including downside risk, credit risk, and liquidity risk. Gao et al. (2020) find media coverage is negatively associated with firms' cost of debt.

Finally, this paper is related to the rising literature on applied machine learning in empirical asset pricing. Gu et al. (2020) forecast individual equity returns with various machine learning algorithms, and Feng et al. (2020) find different portfolio benchmark returns are predictable, and they adopt a forecast combination for individual stock returns. Bianchi et al. (2020) and Feng et al. (2020) applied machine learning and deep learning to investigate the return predictability of U.S. treasury bond returns through revised and real-time macroeconomic variables.

2 Predictive Modeling

As in the empirical asset pricing literature, we describe an asset's excess return as an additive prediction error model:

$$r_{i,t+1} = E_t(r_{i,t+1}) + \epsilon_{i,t+1}. \tag{1}$$

Bonds are indexed as $i = 1, \cdots, N$ and months by $t = 1, \cdots, T_i$. For corporate bond returns, the data structure is extremely imbalanced, and many bonds do not have many observations. Therefore, we have to make the trade-off for predictability heterogeneity and estimation accuracy. We simply assume the same conditional expectation functional form for all bond returns.

We try to model the dynamics of individual corporate bond excess returns, using a large set of predictors in equation 2. The predictors include bond characteristics $z_{i,t}$ and macro-economic variables x_t. The function $g(\cdot)$ can take any functional form. We consider a list of econometric and machine learning methods to approximate the function $g(\cdot)$. Parameters of $g_t(\cdot)$ can be updated by a rolling-window model training scheme:

$$E_t(r_{i,t+1}) = g_t(z_{i,t}, x_t). \tag{2}$$

The most commonly used model in the asset pricing literature is the predictive regression in equation 3. For fitting a pooled regression, the number of observations is much higher than the number of predictors. Therefore, overfitting is not a particular concern, but the bias-variance tradeoff still needs to be considered for out-of-sample forecasting. Also, the nonlinear signals and predictor interactions are ignored in the predictive regression. Gu et al. (2020) and Bianchi et al. (2020) find positive predictability evidence for those methods considering nonlinear signals and predictor interactions. Thus, we call for new predictive modeling methods and list application methods in the following subsections:

$$g(z_{i,t}, x_t; \theta) = [1, z'_{i,t}, x'_t]\theta. \tag{3}$$

2.1 Forecast Combination

Lin et al. (2014) find the mean and median combination of multiple simple regression predictions can predict corporate bond portfolio returns for both in-sample and out-of-sample cases. They use the combination forecasts, which are produced by simple predictive regressions, one predictor at a time. The general combination forecast of $R_{i,t+1}$ is usually the weighted averages of J individual forecasts:

$$\hat{R}^c_{i,t+1} = \sum_{j=1}^{J} \omega_{i,j,t} \hat{R}_{i,j,t+1}, \tag{4}$$

where $R^c_{i,t+1}$ denotes the combination forecast for the return of asset i at time $t+1$, $\hat{R}_{i,j,t+1}$ denotes the return forecast for asset i at time $t+1$ given by predictor j, and $\omega_{i,j,t}$ is the weight to combine individual forecasts. Lin et al. (2014) find the naive combination forecast, such as mean and median forecasts, are strong and robust benchmark models. The mean combination takes equal weights in ω, and the median combination takes the median of $\{\hat{R}_{i,1,t+1}, \hat{R}_{i,2,t+1}, \cdots, \hat{R}_{i,J,t+1}\}$. We include these robust combination forecasts as the benchmark for the prediction exercise.

2.2 Dimension Reduction

We have added two classic dimension reduction techniques: principal component regression (PCR) and partial least squares (PLS). [1] These two approaches are commonly used methods for latent factor models in empirical asset pricing, including recent top publications, namely, Kelly et al. (2019) and Light et al. (2017). One can use a low-dimension version of linearly transformed predictors to construct predictive model and solve the bias-variance trade-off.

PCR consists of a two-step procedure. In the first step, one combines predictors into a small set of linear combinations that best preserve the predictors' covariance structure. In the second step, the first K principal components are used in multiple regressions. PLS also consists of a two-step procedure. In the first step, one combines predictors into a small set of linear combinations that best preserve the covariance between predictors and outcome. In the second step, the first K components are used in multiple regressions. In contrast to PCR, the PLS objective seeks K linear combinations of the independent variables with a maximal predictive association with the dependent variable.

2.3 Regularized Predictive Regressions

Regularized linear predictive regressions, such as lasso and ridge,[2] are also commonly used in machine learning finance. By adding the penalty over ordinary linear regression, these two linear

[1] Both PCR and PLS are introduced in Hastie et al. (2009). The number of components K is set to 3 in our application for both PCA and PLS.

[2] Both lasso and ridge regularized regressions are introduced in Hastie et al. (2009).

machine learning methods preserve the interpretability of linear models. Both Feng et al. (2020) and Kozak et al. (2020) show a natural setup of regularized linear regressions in empirical asset pricing. Compared with PCR and PLS, the advantage of these two models is that they preserve the original predictors without transforming them, and thus keep the interpretation. The disadvantage is the lack of exploring nonlinearity and interaction for predictors.

Lasso and ridge share similar loss functions but have different regularization effects. Lasso adds an L_1 norm penalty on the sum of residual squares, whereas Ridge adds an L_2 norm penalty. On the one hand, lasso can shrink regression coefficients of useless predictors to zero to perform variable selection and achieve a sparse model. On the other hand, ridge shrinks the regression coefficients of useless predictors to very small numbers. A tuning parameter controls for the penalty weight. A larger penalty weight implies more shrinkage on the coefficients. We discuss the selection of tuning parameters in section 4.2.

2.4 Ensemble Tree Models

Regression trees are among the most popular machine learning algorithms, given their intelligibility and simplicity. This approach is an alternative to nonlinear regressions to partition the space into smaller regions, where the interactions are also manageable. The ability to explore nonlinearity and the interaction for predictors provides an advantage of the regression tree model over all the above linear methods. However, the complex structure of a tree makes it prone to overfitting. Thus, we adopt two ensemble tree models to obtain relatively reliable forecasts: gradient boosted regression tree and random forest.[3]

The gradient boosted regression tree is an additive model that makes predictions by combining decisions from a sequence of simple trees. One starts with a shallow tree with a simple structure. The second simple tree fits the residuals from the first tree, instead of the original dependent variable. Then, one further recursively fits the third tree with the second tree's residuals. Finally, we have many trees and many different weak forecasts from simple trees. The predictions by individual trees could be weak, but the prediction from adding up all these tree forecasts is strong. The ensemble forecast is the weighted sum of all these weak forecasts with decreasing weights.

The random forest model also takes the average of the forecasts from many different trees

[3] Both random forest and boosted tree models are introduced in Hastie et al. (2009).

and adopts the bootstrap aggregating scheme. One draws B number of bootstrap samples of the data and fit B different trees. One usually chooses a random subset of predictors for each tree and fits a tree model with the bootstrap sample. The predictions by individual trees could be weak, but the average of all B number of forecasts is strong. Averaging the individual forecasts reduces the overfitting in individual bootstrap samples and provides a stable forecast. In addition, the prediction error can be reduced dramatically for the ensemble forecasts using low related bootstrap samples (bootstrap observations and a random subset of predictors). Tuning parameters and model selection for these two ensemble tree methods are discussed in section 4.2.

3 Empirical Data

3.1 Corporate Bond Returns Sample

The corporate bond observations are collected from four databases: the Lehman Brother Fixed Income (LBFI) database, DataStream, the National Association of Insurance Commissioners (NAIC) database, and the Trade Reporting and Compliance Engine (TRACE) database. We combine the data from these sources to get a large individual corporate bond sample. If we find duplicate observations between multiple sources, we keep only one of them by a priority rank. The rank from high priority to low priority is TRACE, NAIC, LBFI, and DataStream. We prefer to use the transaction-based return data (TRACE) more than the return data based on quotes and matrix calculations (LBFI). The TRACE data start in 2002 and have a short history, whereas the LBFI extends our sample to early 1973.

With the five sources, we can compute corporate bond returns and various characteristics. The monthly corporate bond return at time t is calculated as follows:

$$R_t = \frac{(P_t + A_t) + C_t - (P_{t-1} + A_{t-1})}{P_{t-1} + A_{t-1}}, \tag{5}$$

where P_t is the price, A_t is the accrued interest, and C_t is the coupon payment at time t. We adjust the raw returns to excess returns by subtracting the three-month Treasury bill rate. Our sample excludes any bonds with embedded options. We remove any bond observation whose time-to-maturity is less than two years or longer than 30 years. Summary statistics of the corporate bond returns sample are shown in table B.1 and table B.2.

3.2 Bond Predictors

We have 20 corporate bond characteristics for individual bonds and 20 macro predictors, described in Appendix A.1 and Appendix A.2, respectively.

The corporate bond characteristic sample covers four primary categories, the fundamental characteristics (e.g., rating, duration), the return-distributional characteristics (e.g., variance, downside risk), past returns (e.g., short-term reversal, momentum), and covariance with common factors (e.g., beta on TERM, beta on DEF).

The Mergent FISD database provides corporate bond issue information and many characteristics, such as issue size, issue date, maturity date, coupon interest rate, credit rating, and SIC code. Besides the bond characteristics available from Mergent FISD, we calculate more corporate bond characteristics based on past returns.

First, following those return-distribution characteristics proposed in Bai et al. (2019a) and Bai et al. (2019b), we have added downside risk proxied by the 5% VaR, variance, skewness, and excess kurtosis of the returns, which are estimated using the rolling sample of the past 36 months.

Second, we have created five past return related characteristics, including short-term reversal (lagged one-month returns), momentum (lagged two-month to lagged six-month cumulative returns, and lagged two-month to lagged 12-month cumulative returns), and long-term reversal (lagged 13-month to lagged 24-month cumulative returns, and lagged 13-month to lagged 36-month cumulative returns).

Third, following the risk exposure to the corporate bond's risk factors in Fama and French (1993) and Gebhardt et al. (2005), we estimate the multivariate betas of individual bonds regarding a five-factor model, using the rolling sample of the past 36 months. The five-factor model includes the Fama-French three factors (MktRf for the market, SMB for size, HML for value), plus two bond-related factors (term factor and default factor). We also retain the residual variances from the five-factor model regression.

Finally, in addition to 20 corporate bond characteristics, we have added another group of predictors, 20 macro predictors. These macro predictors cover three main categories: the macroeconomic indicators (e.g., CPI index, three-month Treasure bill rate), the corporate bond market variable (e.g., term spread, corporate bond market return), and the equity market variable (e.g., S&P 500 Index Return, S&P 500 Index Earnings-to-Price Ratio).

4 Empirical Design

4.1 Predictive Modeling Design

Our data cover January 1973 through December 2017. Because we need three years of data to calculate the return-based characteristics, the full sample of predictors starts from January 1976. The out-of-sample evaluation period is January 1998 through December 2017. We recursively estimate the models with a rolling window past 20 years of data and annually update it. Specifically, we estimate the model with the past 20 years of data at the end of each year. We fix the model for 12 months of the next year and predict with the monthly updated predictors.

The individual corporate bond returns data are highly unbalanced. Fitting a time series model for each asset is difficult because of too much missing data. The cross-sectional regression is limited to the information in one month and cannot benefit from the long historical sample period. Therefore, we adopt a pool model by aggregating all the individual bonds in the cross section and covering the 20-year training period. Pool modeling is a common approach for return prediction when the panel is unbalanced, as also seen in Gu et al. (2020) and Bali et al. (2020).

4.2 Deterministic Cross-Validation

Many machine learning methods require tuning parameter selection for the models. Following Feng et al. (2020), we adopt a deterministic five-fold cross-validation scheme as in figure 1. The standard randomized cross-validation is originally designed for independent observations. Our deterministic cross-validation design preserves the continuity of time series predictability in consecutive periods.

Specifically, to predict returns in Year K, we first split the past data up to the end of Year $K-1$ into five consecutive intervals as five folds. Then, we train each model using any four of the five folds and validate using the remaining one fold, which results in five sets of validation errors. Finally, we determine the best tuning parameters according to the average of these five sets' validation errors and refit the model using all five data folds.

Considering the predictors' heterogeneity, we standardize the predictors before the cross-validation. In a particular year K, we calculate $\hat{\mu}_K$ and $\hat{\sigma}_K$ as the sample mean and standard deviation of the past 20 years of data for each predictor. Then, we standardize the sample of the past 20 years of data and the next-year K data with the $\hat{\mu}_K$ and $\hat{\sigma}_K$. The test sample predictors are also standardized by the same sample mean and standard deviation, without any forward-looking concern.

4.3 Performance Evaluation Metrics

4.3.1 Out-of-Sample R-Squared

We follow the prediction literature and assess predictive performance for individual excess bond return forecasts, with the out-of-sample R^2_{OOS} as follows:

$$R^2_{OOS} = 1 - \frac{\sum_{i,t}(r_{i,t} - \hat{r}_{i,t})^2}{\sum_{i,t}(r_{i,t} - \bar{r}_{i,t})^2}, \tag{6}$$

where i denotes different bonds and t denotes different periods in the hold-out sample.

This R^2_{OOS} metric pools prediction errors across all assets and periods to assess the reduction in mean squared forecast error relative to a benchmark forecast. Unlike Gu et al. (2020) and Bali et al. (2020) using a naive forecast of zero as the benchmark, we use the moving average excess return of the corresponding credit rating sorted portfolio in the denominator. The credit rating portfolio benchmark is the most important benchmark for corporate bond evaluation. The moving average of the credit rating portfolio return is calculated over the past 20 years, which is the same time range as the training sample window. In the Appendix table B.3 and table B.4, we also show the out-of-sample R^2 over zero prediction. The credit rating sorted portfolio moving average return is a stronger benchmark than zero prediction.

4.3.2 Fama-Macbeth Out-of-Sample R-Squared

The panel data inference procedure of Fama and MacBeth (1973) has been the standard in the empirical asset pricing literature. We also provide a Fama-Macbeth type out-of-sample R-squared

$\bar{R}^2_{OOS}$ for the evaluation. Specifically, we calculate an out-of-sample R^2 for each month t as equation 7, and we get a time series of them $\{R^2_{OOS,t}\}_{t=1}^{T}$.

$$R^2_{OOS,t} = 1 - \frac{\sum_i (r_{i,t} - \hat{r}_{i,t})^2}{\sum_i (r_{i,t} - \bar{r}_{i,t})^2} \tag{7}$$

The Fama-Macbeth design is only possible because we have a large cross section of observations every month. A large monthly cross section allows us to evaluate the cross-sectional return predictability. The Fama-Macbeth out-of-sample R-squared is [4]

$$\bar{R}^2_{OOS} = \frac{1}{T}\sum_{t=1}^{T} R^2_{OOS,t}. \tag{8}$$

Unlike the simple evaluation of R^2_{OOS}, the Fama-Macbeth modeling further allows us to evaluation the statistical significance of $\bar{R}^2_{OOS}$:

$$H_0 : \bar{R}^2_{OOS} \leq 0 \text{ and } H_1 : \bar{R}^2_{OOS} > 0.$$

Though evaluating if the predictive approach outperforms the benchmark is a one-sided Student's t-test, we still provide the common cutoffs in the tables for empirical results. The standard error for the Fama-Macbeth t-statistic is the sample standard error for $\{R^2_{OOS,t}\}_{t=1}^{T}$.

We would also like to evaluate the prediction performance difference for two sets of results, say $\{R^2_{OOS,t;1}\}_{t=1}^{T}$ and $\{R^2_{OOS,t;2}\}_{t=1}^{T}$. For example, we can evaluate the performance difference with or without one predictor to assess the variable importance. We can also evaluate the performance difference between private and public bond returns.

$$H_0 : \bar{R}^2_{OOS;1} = \bar{R}^2_{OOS;2} \text{ and } H_1 : \bar{R}^2_{OOS;1} \neq \bar{R}^2_{OOS;2}$$

We calculate the t-statistic for the average of the time series difference, which takes the difference of the two time series as shown in Equation 9. We can again employ the Fama-Macbeth approach

[4] The values of $\bar{R}^2_{OOS}$ could be quite different from that of R^2_{OOS} for the nonlinear formula calculation. The R^2_{OOS} aggregates information in the whole panel, but is potentially dominated by the periods with high market volatility or the periods with a large number of observations. However, $\bar{R}^2_{OOS}$ allocates equal weight to each periods, thus partially solves the problems of R^2_{OOS}. The main drawback of $\bar{R}^2_{OOS}$ is that it possibly faces spikes in some month, e.g. a negative performance for a month with limited observations does not affect R^2_{OOS} but affects $\bar{R}^2_{OOS}$. Our solution is to trim the $R^2_{OOS,t}$ time-series at 97.5% and 2.5% quantiles, then calculate the $\bar{R}^2_{OOS}$ and t-stat.

for testing the average of difference, because the monthly out-of-sample R-squared differences are relatively independent.

$$R^2_{OOS,t;\Delta} = R^2_{OOS,t;1} - R^2_{OOS,t;2} \tag{9}$$

4.4 Variable Importance

Besides the machine learning methods' predictive performance, we want to identify the critical predictors, which significantly affect the prediction of individual corporate bond returns. Following Gu et al. (2020), we construct a new predictor dataset by setting all values of predictor j to zero and keep the remaining predictors unchanged.[5] Next, we give a new prediction $\tilde{r}_{i,t}^{[j]}$, with the initially fitted model and the new predictor dataset. Also, we have a new measure for predictability $R^{2,[j]}_{OOS,t}$ defined in equation 10:

$$R^{2,[j]}_{OOS,t} = 1 - \frac{\sum_i (r_{i,t} - \tilde{r}^{[j]}_{i,t})^2}{\sum_i (r_{i,t} - \bar{r}_{i,t})^2} \tag{10}$$

$$\bar{R}^{2,[j]}_{OOS} = \frac{1}{T} \sum_{t=1}^{T} R^{2.[j]}_{OOS,t} \tag{11}$$

With the Fama-Macbeth paired difference t-test, one can test the significance of each predictor. The variable importance of the jth predictor at month t is defined in equation 12, measured by the change in out-of-sample R^2. The variable importance of jth predictor is measured by the expection of $\{\mathrm{vi}^{[j]}_t\}^T_{t=1}$ and its Fama-Macbeth t-stat.

$$\mathrm{vi}^{[j]}_t = R^2_{OOS,t} - \bar{R}^{2,[j]}_{OOS,t} \tag{12}$$

5 Empirical Findings

5.1 Predictability Evidence

We find promising predictability evidence for corporate bond returns over the 20-year out-of-sample period from 1998 to 2017. The main results are reports in table 1. The overall corporate bond returns sample is predictable by the traditional mean combination method, the penalized

[5]Comparing the model performance with and without the predictor is supposedly the ideal approach but is computationally intensive for a high-dimensional predictors.

linear regression (lasso and ridge), the dimension reduction method (PLS), and the non-linear tree model (boosted regression tree and random forest).

In Panel A, we report the prediction performance for investment-grade bond, non-investment-grade bond, and the overall sample. For each case, we report both R^2_{OOS} and $\bar{R}^2_{OOS}$, the latter of which can be tested with the Fama-Macbeth approach. Among the listed machine learning models, we find lasso, ridge, and random forest consistently outperform the 5×1 rating portfolio moving average. The random forest is the best performing model, realizing 4.3% $\bar{R}^2_{OOS}$ for IG, 1.5% for NIG, and 3.6% for the overall sample.

Our findings are consistent with those of Bali et al. (2020). In this paper, the random-forest prediction for the overall bond sample gives an R^2_{OOS} number 2.6% (using 5×1 rating portfolio moving average as the benchmark) and 2.8% (using zero as the benchmark, reported in table B.3) However, we find the predictability for the non-investment-grade bond is lower than that for the investment-grade bond, which is different in Bali et al. (2020). One possible reason is that we train the model for a longer period and report the statistics for the last two decades.

In Panel B, we check the robustness of the predictability evidence under corporate bond sub-samples. The overall corporate bonds are sorted into five credit rating buckets, including AAA, AA, A, BBB, and Junk. The junk bucket is also called non-investment-grade bonds (NIG), and the others are called investment-grade bond (IG). Outperforming the benchmark prediction in every sub-sample is not easy. In our studies, lasso, ridge, and random forest show positive $\bar{R}^2_{OOS}$ among all rating-sorted sub-samples. Ridge and random forest have significantly positive Fama-Macbeth t-stat for $\bar{R}^2_{OOS}$.

In Panel C, we report the results for five duration sub-samples, where the bonds are sorted into five buckets. The Dur1 is the shortest duration bucket, and the Dur5 is the longest duration bucket. Random forest is again the best model and produces significantly positive results among all five duration sub-samples. The mean combination from Lin et al. (2014) also consistently outperforms the benchmark prediction. Finally, we find the predictability evidence is more substantial for bonds with longer durations.

Prediction performance for different time series sub-samples are shown in table 2. We divide the out-of-sample period of 1998 to 2017 into three sub-periods: pre-crisis (1998-2007), crisis (2008-2009), and post-crisis (2010-2017). The mean combination, lasso, and random forest models can significantly predict overall corporate bond returns. The predictability of IG bonds is higher than that of NIG bonds in all three periods. These findings show machine learning forecasts of IG bonds are robust among financial turmoil and economic recession. However, since the financial crisis, we find no significant predictability evidence for NIG bonds.

Finally, we have added a robustness check in table B.4 in the appendix, which uses zero as the prediction benchmark as in Bali et al. (2020). Unsurprisingly, the zero prediction is weaker than the 5×1 rating portfolio moving average benchmark.

5.2 Investment Performance

To exploit the bond return predictability through investment performance, we construct a long-short bond portfolio by the predicted values of individual corporate bond returns. We perform a univariate sort for individual bonds and form equal-weighted long-short strategies on a monthly basis. Figure 2 reports the out-of-sample portfolio cumulative returns. We mainly show results for four predictive methods, including mean combination, lasso, partial least square, and random forest.[6] We find clear monotonic spreads for sorted portfolio returns in each of the four subplots, which is evidence of the return predictability. The long-short portfolio return is substantial, though the premium mainly comes from the long part.

In table 3, we report the monthly average returns of sorted portfolios. In each month, we sort the bonds into quintile equal-weight portfolios, based on the out-of-sample returns prediction. In table 4, we report the evaluation metrics of the long-short portfolio, including the average return, the annualized Sharpe ratio, the risk-adjusted alpha (or pricing error), and the alpha t-statistics for different factor models. The columns show three sub-sample cases, including the IG bond universe, the NIG bond universe, and the whole bond universe. The first row (cbmkt) in each panel reports the long-only equal-weighted portfolio in the corresponding bond universe for a benchmark.

Most out-of-sample long-short strategies deliver higher average returns than the market portfolio for all cases. The non-linear ones (boosted regression tree and random forest) outperform the linear ones (lasso, ridge, and PLS). Noticeably, random forest achieves the best performance with a 1.45% monthly average return for IG, 1.85% for NIG, and 1.48% for the overall sample. More-

[6] Mean combination is from the traditional multiple regression prediction method group. Lasso is from the penalized regression group. PLS is from the dimension reduction group. The random forest is from the non-linear model group.

over, random forest produces a 3.16 annualized Sharpe ratio, which is four times larger than that of CBMKT. Finally, though the NIG portfolios have smaller Sharpe ratios than the IG portfolio, the NIG portfolios have higher average returns.

For the risk-adjusted performance, we show alphas with respect to commonly used factor models in the corporate bond literature, including the five-factor model (FF5) from Fama and French (1993) (MKT, SMB, HML, TERM, and DEF) and the four-factor model (BBW) recently proposed in Bai et al. (2019a) (CBMKT, DRF, CRF, and LRF).[7] Also, we propose a three-factor model (CBMKT, TBL, and STR)[8] indicated by predictive variable importance, detailed in section 5.3.

With respect to FF5 in Panel A and BBW in Panel B, the risk-adjusted alphas are significantly positive for most portfolio returns across all three bond universes. In other words, the FF5 and BBW factor models cannot explain the investment performance of the machine learning strategies, which is why we introduce the predictive variable importance factor model (CBMKT, TBL, and STR). In Panel C, we find the predictive three-factor model shrinks the magnitudes of alphas dramatically compared to panels A and B for almost all cases. However, statistically speaking, the predictive three-factor model only slightly decreases the significance relative to the FF5 and BBW models (i.e., mean combination, lasso, and PLS).

5.3 Variable Importance

Besides the promising predictive performance, identifying those economically and statistically important predictors from among a large predictor set is crucial. We apply a classical empirical finance test for machine learning variable importance as discussed in section 4.4. The empirical results of variable importance are reported in figure 3.

In figure 3, the variable importance is ranked across predictors. The predictors are ordered by their average ranks of variable importance t-stat over different models. Each column corresponds to a predictive model. The color gradients within each column indicate the t-stat values, with the significantly positive values in dark color and the significantly negative ones in light color. Surprisingly, these machine learning methods agree on a small set of predictors as significantly important ones, including the market-wide variables and bond characteristics. They are corporate

[7] We download the factors from Jennie Bai's Website, `http://www.jenniebai.com/data.html`. Their four factors are available since July 2004, so Panel B in table 4 only covers the period "200407-201712."

[8] The CBMKT factor is estimated on our sample, and thus is different from CBMKT used in Bai et al. (2019a).

bond market excess return, three-month treasury bill rate, and short-term reversal. Our findings for short-term reversal are consistent with Bai et al. (2019a). The lagged bond market return is probably the most important predictor for corporate bond returns, which acts as a short-term reversal for the bond market. The three-month treasury bill rate (TBL) is a proxy for the short-term interest rate risk and serves as the risk-free rate to calculate excess returns. In Welch and Goyal (2008), this three-month treasury bill rate is also a positive predictor for the S&P 500 index return.

We find these three important predictors predict future returns and help price the cross section of corporate bond returns. Based on the variable importance, we propose a three-factor model including CBMKT, TBL, and STR.[9] Noticeably, the variable importance results are further supported in table 4. The proposed three-factor model shows pricing performance superior to the five-factor model Fama and French (1993) and the four-factor model in Bai et al. (2019a) with less significant pricing errors.

5.4 Public vs. Private Bonds

Two recently published papers, Chordia et al. (2017) and Choi and Kim (2018), find corporate bond return predictability using equity characteristics. However, Bali et al. (2020) find the predictive power of equity characteristics is redundant when including bond characteristics through an investment approach.[10]. We want to point out one important empirical fact about our corporate bond return sample: observations issued by public companies cover less than 25%. Most previous research involved using equity characteristics drop private bond observations and only use public bond observations.

We investigate this question from a different angle. The efficient market hypothesis supports equal predictability between public and private bonds in the presence of any predictability. If equity characteristics provide additional signals over bond characteristics to return predictability, dropping these predictors must decrease the predictability of public bond returns. Therefore, without equity characteristics, the predictability of public bonds should be smaller than that of private

[9]CBMKT factor is the equal-weighted market portfolio excess return based on our sample. The TBL factor is the three-month treasury bill rate, proxy for short-term interest rate movements. STR is the quintile long-short portfolio return sorted on short-term reversal.

[10]They consider two long-short strategies. One uses the common predictor set of equity and bond characteristics, and the other only uses bond characteristics. They find the portfolio average return difference is insignificant by a Fama-Macbeth t-test.

bonds. We test the below hypothesis:

$$H_0 : \bar{R}^2_{OOS;Private} = \bar{R}^2_{OOS;Public}.$$

In table 5, we evaluate the performance for public and private bonds separately with the same training model. We report the average of the difference time series between the $R^2_{OOS,t}$ of public and private bonds with the Fama-Macbeth two-sample t-statistics. A noticeable finding is that private bonds do have larger $\bar{R}^2_{OOS}$ than public bonds for different models in the whole sample analysis (Panel A). The difference is only significantly positive for mean (median) combination forecast. However, these positive differences are almost insignificant for all sub-period cases (Panel B to D). Therefore, we don't have enough empirical findings, which supports the rejection of the null hypothesis. Consistent with Bali et al. (2020), we find that dropping equity characteristics does not cause significantly lower return predictability for public bonds than for private bonds.

For the cross-sectional sub-sample analysis, we find the indifference of the return predictability between public and private bond is robust to both IG and NIG bonds. In Panel E the IG sub-sample, the paired difference t-test shows 10% confidence level significance only for pca, and 5% for boosted regression tree. As for the NIG sub-sample in Panel F, the mean (median) combination forecast do better predictions for private bonds than public bonds, while other methods don't. As a whole, the there aren't enough empirical findings to claim the private bond returns are more predictable than the public bond, under the IG sample or NIG sample.

5.5 Equity Characteristics and Public Bond Predictability

We can investigate the incremental prediction power of equity characteristics in another way. Putting the commonly used equity characteristics into our predictor set, we further explore the return predictability of the bonds issued by public firms. [11] The main drawback of this approach is that we lost the private bond observations, which takes more than 75% of our sample, because the equity characteristics of private bonds are inaccessible.

In table 6, we report the Fama-Macbeth $\bar{R}^2_{OOS}$ and R^2_{OOS} for the two predictor set specifications. The specification "with" covers 60 predictors, including the 20 equity characteristics, 20 corporate

[11] The equity characteristics are listed in appendix A.3.

bond characteristics, and 20 macro indicators. The specification "without" only covers 40 predictors, excluding the equity characteristics. Also, we have a Fama-Macbeth two sample t-test for the out-of-sample R^2 time series, as reported in column $\bar{R}^2_{OOS;\Delta}$. In Panel A, B, C, and D, we report the results for different sub-periods, and in Panel E and F for different cross-sectional sub-samples.

If the equity characteristics have incremental predictive power against the original 40 predictors, we would expect to see the "with" specification has higher $\bar{R}^2_{OOS}$, significantly positive difference in t-test, and higher R^2_{OOS}. However, we find the results are mixed and messy. In Panel A, which is the full sample results, we find the equity characteristics help improve the $\bar{R}^2_{OOS}$ and R^2_{OOS} for mean (median) combination forecast, boosted regression, and random forest. The improvement is statistically significant only for the mean (median) combination forecast. As for the lasso, ridge, pca, and pls, the inclusion of equity characteristics even deteriorates the prediction results.

For other sub-period results, we find the inclusion of equity characteristics improves the prediction results significantly at 1% confidence level, only in 2010 to 2017 post-financial crisis period and limited for mean (median) combination forecast. For the cross-section sub-sample results, we find the inclusion of equity characteristics improves the prediction results significantly only for Non-investment Grade bonds and is limited for mean (median) combination forecast.

Although the recent literature (see Chordia et al. (2017) and Chordia et al. (2017)) find the equity characteristics have cross-sectional predictive power for corporate bond returns, we don't find empirical evidence for the incremental predictability against the corporate bond characteristics and macro predictors. So, we would not recommend to include equity characteristics in corporate bond returns prediction. The practice of adding equity characteristics is faced with the risk of overfitting and potentially deteriorates the predictability.

5.6 Decomposition of Corporate Bond Returns

Corporate bond returns can be affected by changes in discount rate and cash flow. The cash flow component links to the changes in treasury bond yield curve, while the discount rate component is associated with default risk, liquidity risk, and so on. The previous sections tell that the excess return can be predicted. However, it is still untouched which component is predictable.

To understand the source of corporate bond return predictability, we decompose the excess bond returns into the discount rates and cash flow components. We construct the equivalent Treasury bond portfolio for each bond-month observation. The equivalent Treasury bond portfolio prices are estimated via discounting the expected coupon payments by the corresponding zero-coupon constant maturity Treasury rate. [12] The equivalent bond return (EBR) is calculated from the price changes of the equivalent Treasury bond portfolio.

$$R_{i,t} - R_{f,t} = (R_{i,t} - \text{EBR}_{i,t}) + (\text{EBR}_{i,t} - R_{f,t}) \tag{13}$$

where $R_{i,t}$ is the bond-month return observation, $R_{f,t}$ is the risk free rate proxied by 3-month treasury bill rate, $(R_{i,t} - \text{EBR}_{i,t})$ is the discount rate component, and $(\text{EBR}_{i,t} - R_{f,t})$ is the cash flow component.

Using the 20 bond characteristics and 20 macro predictors, we investigate the two components predictability. To predict the discount rate component, we call the formulation in equation 1 an equation 2, holding the right-hand side predictors, while plug-in the discount rate component on the left-hand side. In table 7, we find the discount rate component is predictable if we look at the whole sample. This predictability evidence is concentrated on Investment-Grade bonds. The Fama-Macbeth $\bar{R}^2_{OOS}$ is significantly positive, and the R^2_{OOS} is always positive for all sub-periods, except the 2008 to 2009 financial crisis. The Non-investment Grade Bonds' discount rate component is relatively harder to predict, as both positive and negative numbers show up in the R^2_{OOS} and the Fama-Macbeth $\bar{R}^2_{OOS}$ is indifferent from zero for many cases. These findings on discount rate are close to the predictability evidence of excess returns in section 5.1. [13] So, the discount rate is an important source of the predictability on excess returns.

According to table 8, it is very difficult to predict the cash flow component, at least compared to the historical average. Almost all methods fail to deliver positive R^2_{OOS} and Fama-Macbeth $\bar{R}^2_{OOS}$ in all sub-periods. [14] Predicting the cash flow component of the corporate bond excess return is es-

[12]The data of zero coupon constant maturity Treasury rate is available from https://fred.stlouisfed.org/ and https://www.federalreserve.gov/data/nominal-yield-curve.htm.

[13]The benchmark (denominator in the formula) in the R^2_{OOS} of the discount rate component is the corresponding historical average of quintile rating portfolio discount rate component, keeping consistent with the excess returns prediction. Specifically, we sort the individual bonds into quintiles based on credit ratings. For each portfolio, we calculate the portfolio discount rate component as the equal-weight average among the underlying assets. Then we check which quintile it belongs to, and use the belonging portfolio historical average as a benchmark prediction for individual asset's discount rate component. This benchmark is empirically stronger than a naive zero prediction.

[14]The benchmark (denominator in the formula) in the R^2_{OOS} of the cash flow component is the corresponding historical average of quintile rating portfolio cash flow component, keeping consistent with the excess returns prediction.

sentially predicting the EBR, which is a portfolio of treasury bonds matching the cash flow payment of the corporate bonds. There are indeed some evidences of treasury bond returns predictability, see Bianchi et al. (2020), Fulop et al. (2020) and Feng et al. (2020). But, the reality is predicting the EBR is more complicated because different corporate bonds have different payment structure and duration. We find it is difficult to predict the cash flow component of corporate bond returns because of either the difficulty in predicting the treasury bond returns or the complex payment structure of different bonds in the cross-section.

6 Conclusion

Our work shows robust positive findings for adopting machine learning methods in the corporate bond literature. The results are in line with recent academic work on machine learning in asset pricing. Linear regularization and dimension-reduction methods produce positive return prediction performance over the benchmark portfolio averages. However, many predictor-return dynamics are undoubtedly nonlinear. We find nonlinear methods, such as boosted regression tree and random forecast, produce robust performance across bond sub-samples and sub-periods and outperform most linear methods. These promising bond return predictability facts via machine learning are important to academic and practitioner researchers. The prediction implied long-short strategy shows additional signals beyond common benchmark factors in equities and corporate bond returns. Further, we find the equity characteristics don't have incremental prediction power against the commonly used corporate bond characteristics and macro predictors. The primary source of corporate bond returns predictability is the discount rate component instead of the cash flow component.

One future direction for extending the bond return predictability study focuses on corporate bonds' portfolio returns. First, like individual equity returns, individual corporate bond returns are extremely noisy due to time-varying strong idiosyncratic shocks. Second, similar to equity portfolio returns, bond portfolio returns might be relatively robust to bond characteristics. Third, creating a portfolio relieves the messy missing data problem due to illiquid bond tradings. In particular, Lin et al. (2014) document return predictability evidence on credit rating or duration sorted bond portfolio returns. Therefore, studying portfolio return predictability with many other bond characteristics could further link pricing sources' connection for corporate bonds.

References

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

Bai, J., T. G. Bali, and Q. Wen (2019b). Do the distributional characteristics of corporate bonds predict their future returns? Technical report, Georgetown University.

Bali, T. G., D. Huang, F. Jiang, and Q. Wen (2020). The cross-sectional pricing of corporate bond using big data and machine learning. Technical report, Georgetown University.

Bianchi, D., M. Büchner, and A. Tamoni (2020). Bond risk premiums with machine learning. *The Review of Financial Studies, Forthcoming*.

Bredendiek, M., G. Ottonello, and R. I. Valkanov (2019). Corporate bond portfolios and asset-specific information. Technical report, Vienna Graduate School of Finance.

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

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

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

Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *American Economic Review 95*(1), 138–160.

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

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Feng, G., A. Fulap, and J. Li (2020). Real-time macro information and bond return predictability: Does deep learning help? Technical report, City University of Hong Kong.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance 75*(3), 1327–1370.

Feng, G., J. He, X. He, and N. G. Polson (2020). Deep learning for predicting asset returns. Technical report, City University of Hong Kong.

Fulop, A., J. Li, and R. Wan (2020). Real-time learning and bond return predictability. *The Journal of Econometrics, Forthcoming*.

Gao, H., J. Wang, Y. Wang, C. Wu, and X. Dong (2020). Media coverage and the cost of debt. *Journal of Financial and Quantitative Analysis 55*(2), 429–471.

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

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Guo, X., H. Lin, C. Wu, and G. Zhou (2020). Extracting information from the corporate yield curve: A machine learning approach. Technical report, Hunan University.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007). The us treasury yield curve: 1961 to the present. *Journal of monetary Economics 54*(8), 2291–2304.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

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

Jostova, G., S. Nikolova, A. Philipov, and C. W. Stahel (2013). Momentum in corporate bond returns. *The Review of Financial Studies 26*(7), 1649–1693.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Light, N., D. Maslov, and O. Rytchkov (2017). Aggregation of information about the cross section of stock returns: A latent variable approach. *The Review of Financial Studies 30*(4), 1339–1381.

Lin, H., J. Wang, and C. Wu (2011). Liquidity risk and expected corporate bond returns. *Journal of Financial Economics 99*(3), 628–650.

Lin, H., J. Wang, and C. Wu (2014). Predictions of corporate bond excess returns. *Journal of Financial Markets 21*, 123–152.

Lin, H., C. Wu, and G. Zhou (2018). Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. *Management Science 64*(9), 4218–4238.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies 21*(4), 1455–1508.

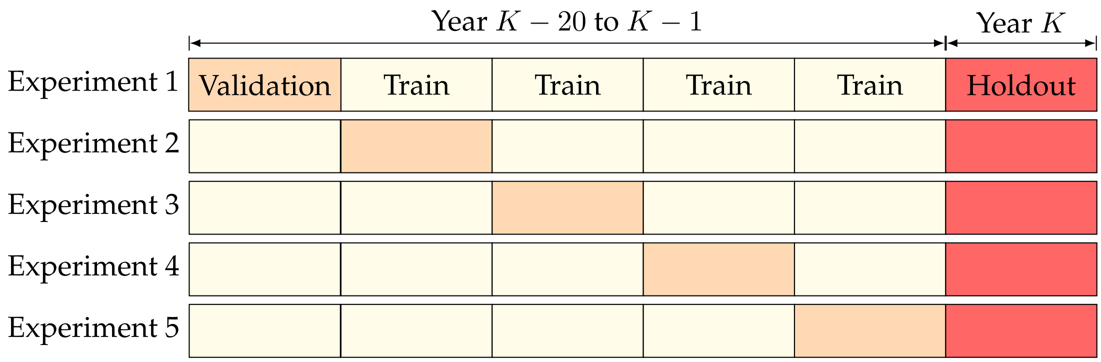

Figure 1: Deterministic Five-Fold Cross-Validation

This figure demonstrates the deterministic five-fold cross-validation scheme. At the end of each year, we re-estimate the models using data for the past 20 years. The standard cross-validation randomly splits data into folds, but our deterministic design divides the sample into five consecutive parts. The rest of the cross-validation procedures follow the standard approach.

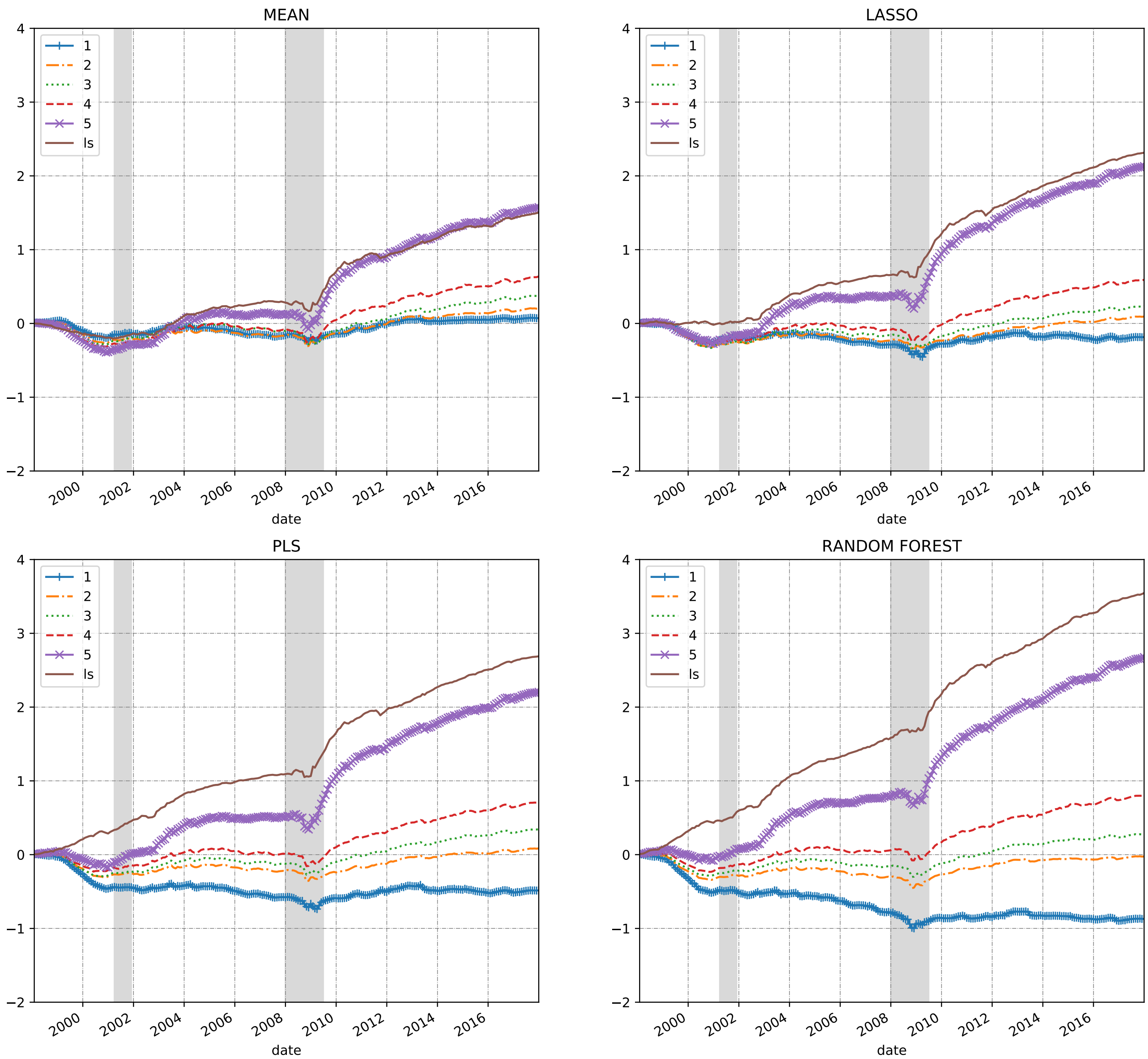

Figure 2: Investment Performance Implied by Bond Return Predictability
This figure reports the investment performance implied by the bond return predictability. We construct quintile sorted portfolios in each month for each prediction method, based on the predicted returns. We also build a long-short strategy by longing the top return forecast portfolio and shorting the bottom return forecast portfolio. The shadow areas are recession periods.

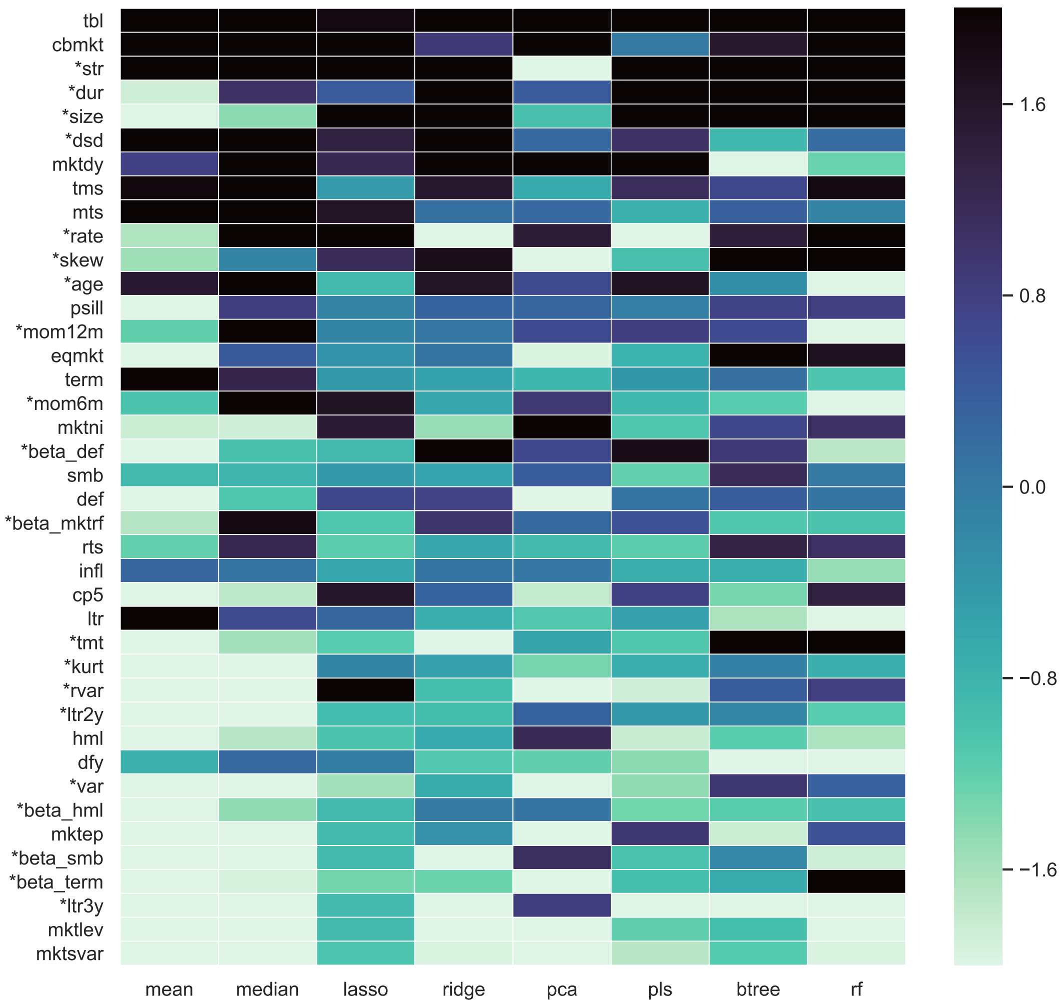

Figure 3: Heat-mapping the t-stat of Variable Importance

Forty predictors are ranked in terms of overall model contribution. Predictors are ordered based on the average rank of variable importance t-stat across all models, with the most influential characteristics on the top and the least significant on the bottom. Columns correspond to individual models, and color gradients within each column indicate the most significant (dark) to least significant (light) variables. The color bar on the right-hand side is within $[-2, 2]$ for t-stat. The characteristics' names are labeled with a * in the vertical axis, whereas the macroeconomic variable names are not labeled.

Table 1: Predicting Individual Corporate Bond Excess Returns

The left panel of this table presents the time series average $\bar{R}^2_{OOS}$, and the corresponding Fama-Macbeth t-statistics as defined in equation 7. On the right panel, we report the out-of-sample R^2_{OOS} (%) of individual corporate bond excess returns as defined in equation 6. The out-of-sample R^2_{OOS} benchmark (denominator) is the moving average excess return for the bond-corresponding 5×1 credit rating portfolio. Panel A reports the results for investment grade, non-investment grade, and the overall bond sample. Panel B reports the results for each sub-sample of corporate bonds sorted by credit rating. Panel C reports the results for each sub-sample of corporate bonds sorted by duration length. We list the performance for eight predictive methods. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$\bar{R}^2_{OOS}$ %						R^2_{OOS} %				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: IG/NIG sub-sample											
	IG	NIG	All				IG	NIG	All		
mean	0.54***	0.15	0.40***				0.39	0.03	0.24		
median	0.11	-0.11	0.03				-0.07	-0.16	-0.11		
lasso	3.27***	1.88***	2.88***				3.75	2.27	3.14		
ridge	3.20**	1.67***	2.76***				3.84	1.37	2.82		
pca	0.77	0.92**	0.89**				0.73	-0.13	0.37		
pls	2.42***	0.80	2.06***				2.89	1.09	2.15		
btree	4.96***	1.08	3.59***				4.65	0.26	2.84		
rf	4.55***	1.85***	3.83***				3.63	1.11	2.59		
Panel B: Rating sub-sample											
	AAA	AA	A	BBB	Junk		AAA	AA	A	BBB	Junk
mean	0.12	0.17	0.57***	0.61***	0.15		0.21	0.02	0.39	0.55	0.03
median	-0.24	-0.26	0.15	0.22**	-0.11		-0.36	-0.51	-0.08	0.17	-0.16
lasso	2.67***	1.62*	2.98***	3.39***	1.88***		2.75	2.86	3.99	4.12	2.27
ridge	2.91***	1.15	2.81***	3.33***	1.67***		4.08	2.80	4.15	3.74	1.37
pca	0.94**	-1.06	0.11	0.94*	0.92**		2.85	0.00	0.51	0.37	-0.13
pls	1.59	-1.51	1.85**	2.76***	0.80		0.98	0.03	3.65	3.61	1.09
btree	7.68***	2.49**	4.31***	4.21***	1.08		7.87	3.63	4.64	3.75	0.26
rf	6.23***	3.76***	4.24***	3.91***	1.85***		6.54	3.42	3.47	2.76	1.11
Panel C: Duration sub-sample											
	Dur1	Dur2	Dur3	Dur4	Dur5		Dur1	Dur2	Dur3	Dur4	Dur5
mean	0.51***	0.41***	0.41***	0.36***	0.39***		0.18	0.16	0.25	0.30	0.25
median	0.00	-0.04	-0.03	0.01	0.07		-0.19	-0.24	-0.18	-0.06	-0.04
lasso	1.28*	1.95***	2.94***	3.03***	3.11***		2.51	2.73	3.60	3.37	3.14
ridge	0.69	1.90**	2.85***	3.13***	3.11***		1.66	2.40	3.47	3.14	2.87
pca	-0.28	0.48	1.06**	0.86*	1.22***		-0.59	0.04	0.70	0.59	0.53
pls	-3.42	-0.37	2.06**	2.71***	3.12***		-0.86	1.03	3.01	2.60	2.90
btree	1.64*	3.06***	3.74***	3.35***	3.57***		2.89	3.88	4.01	2.36	2.16
rf	4.00***	3.88***	3.80***	3.74***	3.81***		2.26	2.60	3.11	2.42	2.56

Table 2: Predicting Individual Corporate Bond Excess Returns
sub-period Analysis

The left panel of This table presents the time series average $\bar{R}^2_{OOS}$, and the corresponding Fama-Macbeth t-statistics as defined in equation 7. On the right panel, we report the out-of-sample R^2_{OOS} (%) of individual corporate bond excess returns as defined in equation 6. The out-of-sample R^2_{OOS} benchmark (denominator) is the moving average excess return for the bond-corresponding 5×1 credit rating portfolio. The results for the overall sample are reported in table 1. This table further reports the results for the sub-periods 1998-2007 (pre-crisis), 2008-2009 (financial crisis), and 2010-2017 (post-crisis). We list the performance for eight predictive methods. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$\bar{R}^2_{OOS}$ %			R^2_{OOS} %		
	IG	NIG	All	IG	NIG	All
Panel A: 1998 - 2010						
mean	0.28	0.73***	0.40**	0.06	-0.11	-0.01
median	-0.07	0.62***	0.11	-0.27	-0.20	-0.24
lasso	2.58***	1.63***	2.37***	2.32	1.35	1.93
ridge	2.85***	1.48***	2.70***	2.75	0.92	2.00
pca	-0.03	0.31	0.46	-0.43	-0.81	-0.58
pls	2.53***	0.99	2.29***	2.59	1.11	1.99
btree	4.16***	0.46	2.64***	3.63	0.00	2.16
rf	3.52***	1.02*	2.81***	3.18	0.80	2.22
Panel B: 2008 - 2009						
mean	0.45***	0.30***	0.35***	0.66	0.28	0.48
median	-0.04	0.07	0.00	0.02	0.02	0.02
lasso	4.88***	3.08***	3.70***	5.71	3.35	4.58
ridge	4.58***	2.16	3.13**	5.34	1.54	3.52
pca	0.63	0.54	0.66	1.84	0.44	1.17
pls	3.38***	2.10	2.19	2.85	0.81	1.87
btree	6.25***	2.74*	4.61***	5.09	0.36	2.83
rf	4.28***	1.87*	3.18***	2.50	1.03	1.79
Panel C: 2010 - 2017						
mean	0.83***	-0.61	0.45***	0.93	-0.25	0.60
median	0.37***	-1.13	0.00	0.37	-0.78	0.05
lasso	3.69***	1.31	2.92***	4.69	2.74	4.16
ridge	3.22***	1.29	2.67***	4.60	3.17	4.21
pca	1.76**	1.42	1.47**	2.34	1.23	2.04
pls	2.17*	0.05	1.61	3.88	2.23	3.43
btree	5.67***	1.50	4.58***	7.04	1.31	5.47
rf	6.03***	2.75**	5.36***	7.02	3.24	5.98

Table 3: Out-of-sample Sorted Portfolio Returns %

This table reports the monthly expected returns of quintile sorted portfolios and long-short portfolios based on the predictions. For each predictive model, we sorted the bonds in the cross-section based on the out-of-sample return predictions. Panel A reports for the overall sample, Panel B restricts the asset pool to investment-grade bonds, and Panel C restricts the asset pool to the non-investment grade bonds. For t-statistics, *, **,and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Low	2	3	4	High	High-Low
			Panel A: ALL			
mean	0.03	0.08	0.16	0.27	0.66	0.63***
median	0.06	0.12	0.20	0.31	0.49	0.43***
lasso	-0.08	0.04	0.10	0.25	0.89	0.96***
ridge	-0.20	0.04	0.14	0.31	0.90	1.09***
pca	0.32	0.20	0.26	0.27	0.14	-0.18*
pls	-0.20	0.04	0.14	0.30	0.92	1.12***
btree	-0.45	0.22	0.22	0.36	0.91	1.36***
rf	-0.36	-0.01	0.12	0.34	1.11	1.48***
			Panel B: IG			
mean	0.00	0.07	0.14	0.22	0.58	0.59***
median	0.02	0.10	0.18	0.28	0.42	0.40***
lasso	-0.12	0.03	0.08	0.21	0.80	0.91***
ridge	-0.24	0.01	0.14	0.27	0.83	1.06***
pca	0.32	0.18	0.24	0.26	0.01	-0.30***
pls	-0.25	0.01	0.13	0.27	0.84	1.08***
btree	-0.54	0.23	0.22	0.34	0.84	1.38***
rf	-0.40	-0.05	0.10	0.30	1.05	1.45***
			Panel C: NIG			
mean	0.14	0.17	0.24	0.53	1.03	0.90***
median	0.36	0.32	0.30	0.43	0.65	0.29
lasso	-0.12	0.07	0.16	0.64	1.34	1.47***
ridge	-0.18	0.10	0.25	0.53	1.39	1.57***
pca	0.83	0.31	0.32	0.29	0.36	-0.46*
pls	-0.20	0.07	0.24	0.47	1.52	1.71***
btree	-0.33	0.24	0.43	0.49	1.38	1.74***
rf	-0.40	0.13	0.38	0.53	1.45	1.85***

Table 4: Prediction Implied Long-Short Strategy Returns

This table reports the summary statistics for out-of-sample long-short portfolio returns. The statistics include monthly expected returns (%), alphas (%) on a factor model, t test result for the alpha, and annualized Sharpe ratios. In Panel A, we test the portfolios with the five-factor model (MKT, SMB, HML, TERM, DEF) in Fama and French (1993) with test period from January 1998 to December 2017. In Panel B, the factor model is the variable importance implied factor model (CBMKT, TBL, and short-term reversal factor) and the test period matches Panel A. In Panel C, the factor model follows Bai et al. (2019a) (CBMKT, DRF, CRF, and LRF), with test period from July 2004 to December 2017, because the data are available since July 2004. In Panel D, the factors copy Panel B and the test period copies Panel C. For t-statistics, *, **,and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	IG			NIG			All		
	Mean	α	SR	Mean	α	SR	Mean	α	SR
Panel A: FF5 199801-201712									
cbmkt	0.23	0.07	0.66	0.49	0.26*	0.61	0.27	0.10*	0.73
mean	0.59	0.50***	1.48	0.90	0.91***	0.74	0.63	0.56***	1.25
median	0.40	0.32***	1.23	0.29	0.24	0.34	0.43	0.36***	1.02
lasso	0.91	0.85***	2.17	1.47	1.42***	1.21	0.96	0.90***	1.94
ridge	1.06	1.03***	2.67	1.57	1.55***	1.27	1.09	1.05***	2.22
pca	-0.30	-0.39***	-0.77	-0.46	-0.56**	-0.43	-0.18	-0.28***	-0.39
pls	1.08	1.05***	2.74	1.71	1.70***	1.35	1.12	1.08***	2.30
btree	1.38	1.39***	3.56	1.74	1.74***	1.72	1.36	1.40***	3.04
rf	1.45	1.37***	3.90	1.85	1.72***	1.74	1.48	1.40***	3.16
Panel B: BBW 200407-201712									
cbmkt	0.37	0.07**	1.07	0.72	0.04	0.84	0.42	0.07	1.12
mean	0.79	0.39***	1.78	1.33	1.01***	1.07	0.84	0.43***	1.53
median	0.57	0.26***	1.62	0.74	0.19	0.95	0.65	0.24**	1.42
lasso	1.13	0.82***	2.50	1.67	1.41***	1.34	1.17	0.87***	2.21
ridge	1.05	0.72***	2.28	1.62	1.27***	1.32	1.10	0.76***	2.00
pca	-0.05	-0.12	-0.12	-0.02	-0.29	-0.02	0.12	-0.04	0.24
pls	1.07	0.76***	2.35	1.62	1.35***	1.30	1.12	0.79***	2.07
btree	1.39	1.11***	3.25	1.47	1.16***	1.60	1.35	1.07***	2.75
rf	1.44	1.12***	3.56	1.93	1.20***	1.98	1.49	1.09***	2.96
Panel C: (CBMKT TBL STR) 199801-201712									
cbmkt	0.23	0.04	0.66	0.49	-0.30	0.61	0.27	0.00	0.73
mean	0.59	0.44***	1.48	0.90	0.01	0.74	0.63	0.24	1.25
median	0.40	0.55***	1.23	0.29	0.68**	0.34	0.43	0.49***	1.02
lasso	0.91	0.47***	2.17	1.47	-0.01	1.21	0.96	0.24*	1.94
ridge	1.06	0.48***	2.67	1.57	-0.11	1.27	1.09	0.23	2.22
pca	-0.30	0.37***	-0.77	-0.46	1.89***	-0.43	-0.18	0.79***	-0.39
pls	1.08	0.46***	2.74	1.71	-0.23	1.35	1.12	0.22	2.30
btree	1.38	1.08***	3.56	1.74	0.64*	1.72	1.36	0.97***	3.04
rf	1.45	1.02***	3.90	1.85	0.65*	1.74	1.48	0.98***	3.16

Table 5: Predicting Private and Public Bonds

This table reports the time series average $\bar{R}^2_{OOS}$ (%) and the corresponding Fama-Macbeth t-statistics as defined in equation 7. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The abbreviation "Pvt" is for Private bond, and "Plc" is for Public bond. For the public-private bond predictability difference, we report "t-test" as the Fama-Macbeth t-stat of the difference time series of $\bar{R}^2_{OOS;\Delta}$ defined in Equation 9. This table also presents the overall R^2_{OOS} (%) of individual corporate bond excess returns as defined in equation 6. The out-of-sample R^2_{OOS} benchmark (denominator) is the moving average excess return for the bond-corresponding 5×1 credit rating portfolio. Panel A reports the results for private company bonds, public company bonds, and their predictability difference. The Panel B, C, and D report the results for the sub-periods 1998-2007 (pre-crisis), 2008-2009 (financial crisis), and 2010-2017 (post-crisis). The Panel E and F report the results for the sub-samples: investment-grade and non-investment-grade bonds.

	$\bar{R}^2_{OOS}$		$\bar{R}^2_{OOS;\Delta}$	R^2_{OOS}		$\bar{R}^2_{OOS}$		$\bar{R}^2_{OOS;\Delta}$	R^2_{OOS}	
	Pvt	Plc	t-test	Pvt	Plc	Pvt	Plc	t-test	Pvt	Plc
	Panel A: ALL 1998-2017					Panel B: ALL 1998-2007				
mean	0.41***	0.37***	0.33*	0.22	0.31	0.40**	0.39**	0.03	-0.03	0.06
median	0.03	0.04	0.38*	-0.13	-0.03	0.09	0.15	-0.01	-0.27	-0.13
lasso	2.99***	2.35***	0.11	3.30	2.58	2.51***	1.96***	0.49	2.05	1.52
ridge	2.88***	2.27***	0.22	2.73	2.79	2.78***	2.22***	0.45	2.09	1.69
pca	0.89**	0.69	-0.15	0.28	0.71	0.46	0.06	0.37	-0.60	-0.54
pls	2.17***	1.22	0.49	2.20	1.98	2.28***	1.68*	0.56	2.08	1.66
btree	3.84***	2.60***	-0.23	3.25	1.44	3.04***	1.50	1.84**	2.61	0.62
rf	3.94***	3.22***	0.89	2.78	1.94	3.00***	2.20***	1.00**	2.54	1.10
	Panel C: ALL 2008-2009					Panel D: ALL 2010-2017				
mean	0.35***	0.34**	-0.03	0.46	0.55	0.46***	0.37**	0.08	0.59	0.64
median	0.00	-0.05	0.03	0.00	0.09	0.00	-0.07	0.07	0.06	0.04
lasso	3.92***	3.67*	0.72	4.89	3.56	2.99***	1.96	1.46	4.16	4.15
ridge	3.03**	3.09	-0.29	3.39	3.96	2.92***	1.43	1.85	4.28	3.95
pca	0.63	0.86	-0.57	0.94	1.95	1.53**	1.35	0.53	1.95	2.36
pls	2.30	1.44	1.12	1.83	2.03	1.85	-0.03	2.07	3.50	3.14
btree	4.84***	3.73**	2.21	3.33	1.17	4.59***	4.06**	2.36	5.46	5.48
rf	3.15***	3.31**	0.09	1.78	1.84	5.36***	4.54***	2.09	6.09	5.58
	Panel E: IG 1998-2017					Panel F: NIG 1998-2017				
mean	0.53***	0.56***	-0.04	0.34	0.56	0.20	-0.11	0.31**	0.03	0.03
median	0.10	0.18	-0.08	-0.11	0.08	-0.06	-0.40*	0.34**	-0.17	-0.16
lasso	3.39***	2.56***	0.83	3.86	3.31	2.07***	1.58**	0.49	2.44	1.76
ridge	3.29***	2.56***	0.73	3.75	4.22	1.82***	1.33*	0.49	1.43	1.18
pca	0.84***	0.14	0.70*	0.58	1.34	0.96**	0.76	0.20	-0.18	0.02
pls	2.51*	1.70	0.80	2.78	3.31	0.86	0.12	0.74	1.30	0.51
btree	5.27***	3.46***	1.81**	5.15	2.71	0.78	1.56	-0.78	0.34	0.05
rf	4.65***	3.95***	0.70	3.74	3.19	2.21***	1.51*	0.69	1.31	0.56

Table 6: Public Bond Predictability with and without Equity Characteristics

This table reports the time series average $\bar{R}^2_{OOS}$ (%) and the corresponding Fama-Macbeth t-statistics as defined in equation 7. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The "with" specification, we include the 20 equity characteristics, and for the "without" specification, we don't add the equity characteristics. For the bond predictability difference the two specifications, we report "t-test" as the Fama-Macbeth t-stat of the difference time series of $\bar{R}^2_{OOS;\Delta}$ defined in Equation 9. This table also presents the overall R^2_{OOS} (%) of individual corporate bond excess returns as defined in equation 6. The out-of-sample R^2_{OOS} benchmark (denominator) is the moving average excess return for the bond-corresponding 5×1 credit rating portfolio. Panel A reports the results for private company bonds and their predictability difference. The Panel B, C, and D report the results for the sub-periods 1998-2007 (pre-crisis), 2008-2009 (financial crisis), and 2010-2017 (post-crisis). The Panel E and F report the results for the sub-samples: investment-grade and non-investment-grade bonds.

	$\bar{R}^2_{OOS}$		$\bar{R}^2_{OOS;\Delta}$	R^2_{OOS}		$\bar{R}^2_{OOS}$		$\bar{R}^2_{OOS;\Delta}$	R^2_{OOS}	
	with	without	t-test	with	without	with	without	t-test	with	without
	Panel A: ALL 1998-2017					Panel B: ALL 1998-2007				
mean	0.57***	0.37***	0.18**	0.28	0.31	0.23	0.39**	-0.17***	-0.15	0.06
median	0.26**	0.04	0.24***	-0.02	-0.03	0.01	0.15	-0.14***	-0.34	-0.13
lasso	2.19***	2.35***	-0.12	2.52	2.58	2.56***	1.96***	0.54**	2.33	1.52
ridge	1.91**	2.27***	-0.48	2.13	2.79	2.58***	2.22***	0.07	2.19	1.69
pca	-0.06	0.69	-0.70**	0.27	0.71	-0.97	0.06	-1.18***	-1.56	-0.54
pls	-1.01	1.22	-2.84***	0.71	1.98	1.76	1.68*	-0.29	2.46	1.66
btree	3.09***	2.60***	0.83	1.67	1.44	2.79***	1.50	1.31**	3.49	0.62
rf	3.02***	3.22***	0.18	2.78	1.94	2.44***	2.20***	0.15	1.85	1.10
	Panel C: ALL 2008-2009					Panel D: ALL 2010-2017				
mean	0.34**	0.34**	-0.03	0.57	0.55	1.06***	0.37**	0.67***	1.13	0.64
median	0.00	-0.05	0.05	0.17	0.09	0.68***	-0.07	0.78***	0.67	0.04
lasso	4.10	3.67*	-0.35	2.61	3.56	0.73	1.96	-0.90	2.94	4.15
ridge	2.29	3.09	-2.16	1.72	3.96	0.42	1.43	-0.75	3.00	3.95
pca	-1.57	0.86	-2.48***	1.95	1.95	1.44	1.35	0.36	2.49	2.36
pls	-0.18	1.44	-3.37	-0.81	2.03	-5.63**	-0.03	-5.91***	-1.64	3.14
btree	-1.12	3.73**	-3.91	-2.06	1.17	4.57***	4.06**	1.40	5.01	5.48
rf	3.32**	3.31**	0.21	3.03	1.84	3.72***	4.54***	0.22	5.57	5.58
	Panel E: IG 1998-2017					Panel F: NIG 1998-2017				
mean	0.57***	0.56***	0.11	0.28	0.56	0.16	-0.11	0.33***	0.00	0.03
median	0.26**	0.18	0.20**	-0.02	0.08	-0.12	-0.40*	0.38***	-0.16	-0.16
lasso	2.19***	2.56***	-0.77	2.52	3.31	2.40***	1.58**	0.71*	1.88	1.76
ridge	1.91**	2.56***	-0.82	2.13	4.22	1.35*	1.33*	-0.01	0.02	1.18
pca	-0.06	0.14*	-1.13***	0.27	1.34	0.27	0.76	-0.41	0.13	0.02
pls	-1.01	1.71	-3.94***	0.71	3.31	-1.33	0.12	-1.81*	-0.35	0.51
btree	3.09***	3.46***	0.26	1.67	2.70	1.91*	1.56	0.55	1.89	0.05
rf	3.02***	3.95***	-0.02	2.78	3.19	1.63**	1.51*	0.51	1.84	0.56

Table 7: Predicting Discount Rate Component

The left panel of This table presents the time series average $\bar{R}^2_{OOS}$, and the corresponding Fama-Macbeth t-statistics as defined in equation 7. On the right panel, we report the out-of-sample R^2_{OOS} (%) (defined in equation 6) of the discount rate component of individual corporate bond returns. The out-of-sample R^2_{OOS} benchmark (denominator) is the moving average discount rate component for the bond-corresponding 5×1 credit rating portfolio. This table further reports the results for different the sub-periods: 1998-2017 (whole oos period) 1998-2007 (pre-crisis), 2008-2009 (financial crisis), and 2010-2017 (post-crisis). We list the performance for eight predictive methods. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$\bar{R}^2_{OOS}$ %			R^2_{OOS} %		
	IG	NIG	All	IG	NIG	All
Panel A: 1998 - 2017						
mean	0.73***	0.09	0.53***	0.43	-0.08	0.21
median	0.38***	-0.15	0.22**	0.03	-0.24	-0.08
lasso	3.39***	1.85***	2.89***	4.15	1.44	2.99
ridge	3.54***	1.92***	3.01***	3.12	0.70	2.08
pca	1.03**	0.47	0.92**	0.56	-0.21	0.23
pls	3.00***	1.11	2.44***	3.36	0.62	2.19
btree	4.93***	0.32	3.47***	1.34	-4.61	-1.20
rf	3.76***	1.53**	3.07***	1.34	-1.10	0.30
Panel B: 1998 - 2010						
mean	0.37**	0.45*	0.37**	0.16	-0.18	0.01
median	0.09	0.35	0.15	-0.14	-0.29	-0.20
lasso	2.32***	1.55***	1.83***	2.62	1.12	1.97
ridge	2.66***	1.42***	2.08***	3.26	1.60	2.54
pca	0.81	0.57	0.81*	0.82	-0.28	0.34
pls	2.21**	0.55	1.47*	3.78	1.87	2.95
btree	2.96***	0.28	1.78**	4.39	1.36	3.08
rf	3.50***	1.28**	2.56***	4.06	1.71	3.05
Panel C: 2008 - 2009						
mean	0.37	0.22*	0.28**	0.47	0.11	0.29
median	-0.16	0.02	-0.08	-0.10	-0.06	-0.08
lasso	6.83***	1.85	4.62***	8.02	2.06	5.12
ridge	3.74**	0.51	2.31	2.91	-0.58	1.21
pca	0.43	0.34	0.30	0.48	0.08	0.28
pls	5.83**	1.60	4.08*	3.81	-0.63	1.65
btree	-3.33	-8.00	-5.97	-5.59	-12.58	-8.99
rf	-0.44	-1.73	-1.39	-2.61	-4.42	-3.49
Panel D: 2010 - 2017						
mean	1.32***	-0.37	0.81***	1.03	-0.33	0.66
median	0.86***	-0.81	0.36***	0.67	-0.76	0.28
lasso	3.72***	1.93	3.34***	1.65	0.53	1.35
ridge	4.61***	2.78**	4.21***	3.10	1.41	2.64
pca	1.37	-0.14	0.92	0.08	-1.07	-0.23
pls	3.23**	1.53	2.94**	1.60	-0.55	1.02
btree	9.39***	2.39	7.71***	5.00	-1.99	3.10
rf	4.93***	2.51	4.45***	1.00	-1.63	0.28

Table 8: Predicting Cash Flow Component

The left panel of This table presents the time series average $\bar{R}^2_{OOS}$, and the corresponding Fama-Macbeth t-statistics as defined in equation 7. On the right panel, we report the out-of-sample R^2_{OOS} (%) (defined in equation 6) of the cash flow component of individual corporate bond returns. The out-of-sample R^2_{OOS} benchmark (denominator) is the moving average cash flow component for the bond-corresponding 5×1 duration portfolio. This table further reports the results for different the sub-periods: 1998-2017 (whole oos period) 1998-2007 (pre-crisis), 2008-2009 (financial crisis), and 2010-2017 (post-crisis). We list the performance for eight predictive methods. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$\bar{R}^2_{OOS}$ %			R^2_{OOS} %		
	IG	NIG	All	IG	NIG	All
			Panel A: 1998 - 2017			
mean	0.39	0.80	0.39	-0.16	-0.23	-0.17
median	0.06	0.41	0.06	-0.26	-0.24	-0.26
lasso	-1.21	-4.00	-1.26	-0.72	-0.79	-0.73
ridge	1.30	-1.98	1.14	-0.20	-0.15	-0.19
pca	-3.28	-5.53	-3.52	-1.50	-2.44	-1.61
pls	-22.26	-59.15	-24.23	-7.76	-11.16	-8.19
btree	-40.46	-59.02	-41.04	-17.12	-22.02	-17.74
rf	-23.59	-44.14	-24.39	-8.80	-12.61	-9.29
			Panel B: 1998 - 2010			
mean	-0.06	0.73	-0.05	-0.20	-0.07	-0.18
median	-0.54	0.14	-0.55	-0.38	-0.23	-0.36
lasso	-3.26	-6.48	-3.27	-0.39	-0.31	-0.38
ridge	-0.73	-4.62	-0.95	-1.88	-1.59	-1.85
pca	-6.17	-6.74	-6.17	-2.54	-2.53	-2.53
pls	-23.77	-62.83	-26.34	-6.92	-7.34	-6.98
btree	-52.92	-76.78	-53.78	-12.41	-12.20	-12.38
rf	-27.78	-50.93	-28.50	-3.26	-4.10	-3.37
			Panel C: 2008 - 2009			
mean	-0.55	-1.66	-0.65	-0.58	-0.78	-0.61
median	-0.33	-0.99	-0.37	-0.18	-0.20	-0.18
lasso	-2.12	-5.46	-2.25	-4.34	-4.01	-4.29
ridge	-0.20	-3.89	-0.43	2.49	2.65	2.51
pca	2.47	-3.03	1.76	-1.09	-2.45	-1.28
pls	-58.33	-90.89	-61.24	-21.96	-31.59	-23.32
btree	-56.56	-70.67	-57.24	-48.68	-55.13	-49.58
rf	-34.63	-51.95	-36.33	-31.84	-40.28	-33.03
			Panel D: 2010 - 2017			
mean	1.03*	1.59	1.05*	0.18	-0.11	0.15
median	0.82	1.28	0.84	-0.08	-0.29	-0.11
lasso	1.34	0.24	1.29	0.93	0.81	0.92
ridge	2.87**	2.20	2.79*	1.19	0.75	1.14
pca	-0.88	-4.04	-1.40	0.14	-2.21	-0.12
pls	-12.09	-40.13	-12.52	-0.44	-2.80	-0.70
btree	-15.85	-29.31	-16.09	-6.04	-16.56	-7.19
rf	-12.96	-27.96	-13.46	-4.55	-8.71	-5.00

Appendices

A Predictor Descriptions

A.1 Corporate Bond Characteristics

A.1.1 Fundamental Characteristics

We have five fundamental characteristics: credit rating (CRT), time-to-maturity (TMT), age (AGE), duration (DUR), and amount outstanding (SIZE).

We collect individual bond level credit ratings (CRT) from Mergent FISD. All ratings are in numerical numbers. For example, 0 refers to a AAA rating, 1 refers to AA+, ..., and 21 refers to D. Investment-grade bonds have ratings from 0 (AAA) to 9 (BBB-). Non-investment-grade bonds have ratings above 9. If both Moody's and S&P ratings are available, we use the average.

TMT is the number of years left for a bond from the current month to the maturity date. AGE is the time between the current trading month and the issuance data, in number of years. DUR is calculated as the Macaulay duration with yield-to-maturity from Gürkaynak et al. (2007). SIZE is collected from Mergent FISD.

A.1.2 Return-based Characteristics

Following Jostova et al. (2013), we have created five past return related characteristics, including STR (lagged one-month returns as short-term reversal), mom6m (lagged two-month to lagged six-month cumulative returns as momentum), mom12m (lagged two-month to lagged 12-month cumulative returns as momentum), ltr2y (lagged 13-month to lagged 24-month cumulative returns as long-term reversal), and ltr3y (lagged 13-month to lagged 36-month cumulative returns as long-term reversal). Following Bai et al. (2019a) and Bai et al. (2019b), we include VAR (variance), DSD (downside risk 5% VaR), SKEW (skewness), and KURT (kurtosis). These four distributional characteristics are calculated using the rolling sample of the past 36 months.

A.1.3 Risk Characteristics

Following Fama and French (1993), we list five important risk factors for corporate bond returns: MKT, SMB, HML, TERM, and DEF. We estimate the multivariate beta of individual bonds on the factors using the rolling sample of the past 36 months. In this part, we have BETA_MKT, BETA_SMB, BETA_HML, BETA_DEF, BETA_TERM, and the corresponding residual variances RVAR.

A.2 Macro Predictors

We have 20 macro predictors in this category, including S&P 500 predictors (dividend-to-price, earnings-to-price, net equity issuance, leverage, and stock variance), Pastor-Stambaugh illiquidity, duration spread, rating spread, corporate bond market returns, Fama-French three factors (MktRf, SMB, and HML), term factor, and default factor, one-month Treasury bill rate, inflation, long-term returns, default yield, term spread, and Cochrane-Piazessi forward factor.

A.2.1 Fama-French Three Factors (MktRf, SMB, HML)

We use the Fama-French three factors (MktRf for excess market return, SMB for size factor, and HML for value factors 3), which are downloaded from French's Website [15].

A.2.2 Corporate Bond Market Factor (CBMKT)

Our corporate bond market factor (cbmkt) is the equal-weighted average return of all observed corporate bond excess returns in each month.

A.2.3 Term Factor (TERM)

Following Fama and French (1993), TERM proxies the unexpected changes in interest rates. It is constructed as the difference between the monthly long-term government bond return (from Ibbotson Associates) and the one-month Treasury bill rate measured at the end of the previous month.

[15] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

A.2.4 Default Factor (DEF)

Following Fama and French (1993), the default factor arises from the shifts in economic conditions that change the likelihood of default. It is proxied by DEF, 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.

A.2.5 3-Month Treasury Bills (TBL)

Following Welch and Goyal (2008), the TBL is the 3-Month Treasury Bill : Secondary Market Rate from the economic research database at the Federal Reserve Bank at St. Louis Website [16].

A.2.6 Inflation (INFL)

Following Welch and Goyal (2008), the INFL is the *Consumer Price Index (All Urban Consumers)* from the Bureau of Labor Statistics.

A.2.7 Long-Term Rate of Returns (LTR)

Following Welch and Goyal (2008), the LTR is the long-term rate of returns provided by Ibbotson Associates.

A.2.8 Term Spread (TMS)

Following Welch and Goyal (2008), the TMS is the difference between the long-term yield on government bonds (by Ibbotson Associates) and the Treasury bill. TMS is different from TERM in A.2.3, because TMS uses the long-term yield on government bonds, whereas TERM uses the long-term return on government bonds.

A.2.9 Default Yield Spread (DFY)

Following Welch and Goyal (2008), the DFY is the difference between BAA- and AAA-rated corporate bond yields.

[16] https://fred.stlouisfed.org/series/DTB3

A.2.10 Cochrane-Piazzesi Forward Factor (CP5)

Following Cochrane and Piazzesi (2005), we construct a five-year forward factor with the codes published on John Cochrane's website [17].

A.2.11 Maturity Spread (MTS)

We construct a short maturity portfolio (two to five years) and a long maturity portfolio (longer than 10 years) in each month. The MTS is the spread between the long and short maturity equal-weighted portfolio returns.

A.2.12 Rating Spread (RTS)

The RTS is the spread between the AAA and Junk rating equal-weighted portfolio returns.

A.2.13 Paster-Stambaugh Illiquidity (PSILL)

Following Pástor and Stambaugh (2003), we download the aggregate illiquidity measure from Robert Stambaugh's Website [18].

A.2.14 Equity Market Aggregate Variables

We download the data in Welch and Goyal (2008) from Amit Goyal's website[19] and calculate the S&P 500 aggregate characteristics, including MKTDY (Dividend Yield), MKTEP (Earnings-to-Price), MKTNI (Net Equity Issuance), and MKTSVAR (Stock Variance).

A.3 Equity Characteristics

Following Hou et al. (2020), we categorize the equity characteristics into six groups. The data are available on their website http://global-q.org/testingportfolios.htm. We replicate their downloadable factors successfully. Then, we follow their equity characteristics formulas and create 20 representative characteristics for our data library.

[17] https://faculty.chicagobooth.edu/john.Cochrane/research/Data_and_Programs/Bond_Risk_Premia/index.htm

[18] http://finance.wharton.upenn.edu/~stambaug/liq_data_1962_2018.txt

[19] http://www.hec.unil.ch/agoyal/.

For our research, we need data that can be updated every month. However, many of these characteristics calculations involve quarterly or even annually updated financial variables. Therefore, we need some adjustments to make them update monthly, and our data updating principle is straightforward. We update those formula variables as late as possible. For example, corporate earnings are updated quarterly from the corporate earnings report, and the share price can be updated monthly (or at a higher frequency). We calculate the earnings-to-price ratio by updating the nominator quarterly and the denominator monthly.

Finally, we take a six-month lag for using the data for accounting information from the annual report. We use their annually updated accounting variables at the end of June in the second year for most companies. And for quarterly accounting information, we take a three-month lag. Though conservative, these procedures are standard for using data in Compustat.

A.3.1 Momentum

Cumulative abnormal returns around earnings announcement dates (ABR)

Standard Unexpected Earnings (SUE)

Revisions in analyst earnings forecasts (RE)

Cumulative Returns on prior 2-12 month (MOM12M)

A.3.2 Value versus growth

Book-to-Market (BM)

Earnings-to-Price (EP)

Cashflow-to-Price (CFP)

Sales-to-Price (SP)

A.3.3 Investment

Asset Growth Rate (AGR)

Net Equity Issuance (NI, or share repurchase)

Accruals (ACC)

A.3.4 Profitability

Operating Profitability (OP)

Return on Equity (ROE)

A.3.5 Intangibles

Seasonality (SEAS1A)

Advertisement-to-Market (ADM)

Research and Design Expense to market (RDM)

A.3.6 Frictions (or Size)

Market Equity (ME)

Stock Variance (SVAR)

CAPM Beta (BETA)

Short-term Reversal (MOM1M)

B Additional Tables

Table B.1: Descriptive Statistics

Our final data sample includes 589,528 monthly return observations of 19,782 unique corporate bonds from January 1976 to December 2017. The raw data starts in the year 1973. However, we require a 3-year window to initialize risk characteristics such as β_{term}, β_{smb}, and so on. We report TRACE and NAIC together because they are both transaction-based data, and a large proportion of NAIC observations is covered by TRACE.

	All Databases	Lehman	DataStream	TRACE&NAIC
Bond-month observations	589,528	126,710	63,107	399,711
Period coverage	1976-2017	1976-1998	1990-2008	1994-2017
Return - mean (%/month)	0.51	0.73	0.50	0.44
Return - median (%/month)	0.52	0.79	0.61	0.52
Excess Return - mean (%/month)	0.14	0.15	0.18	0.13
Excess Return - median (%/month)	0.22	0.27	0.25	0.18
Rating - mean	5.62	5.18	7.21	5.51
Rating - median	5.00	5.00	7.00	5.00
% of rated that are IG	89.64	93.53	84.20	89.27
% of rated that are NIG	10.36	6.47	15.80	10.73
% of Private bond	78.81	82.66	94.37	75.13
Duration - mean (years)	6.25	5.56	8.67	6.09
Duration - median (years)	5.48	5.32	9.00	5.06
Age - mean (years)	8.32	17.61	6.42	5.68
Age - median (year)	5.37	19.90	5.89	3.97
Amt outst. - mean ($ millions)	10,174.41	708.36	1,861.02	11,441.48
Amt outst. - mean ($ millions)	1,500.00	200.00	1,000.00	2,000.00

Table B.2: Individual Corporate Bond Returns Sample Distribution (%)

Our data sample includes 589,528 monthly return observations of 19,782 unique corporate bonds from January 1976 to December 2017. Panel A reports the percentage distribution of the corporate bond data by credit ratings and maturity length. Panel B reports the percentage distribution of the data by rating and data source.

	AAA	AA	A	BBB	Junk	All
Panel A: By Rating & Maturity						
2	1.87	2.66	5.45	2.78	1.11	13.87
3	1.34	2.17	4.62	2.44	1.00	11.56
4	1.34	2.06	4.61	2.39	0.98	11.37
5	0.79	1.27	3.04	1.78	0.87	7.75
6	0.80	1.16	2.92	1.77	0.86	7.51
7	0.60	1.04	2.54	1.54	0.66	6.39
8	0.59	1.03	2.47	1.56	0.59	6.24
9	0.57	1.01	2.57	1.75	0.62	6.51
10	0.16	0.53	1.07	0.88	0.33	2.96
>10	1.97	3.34	8.93	8.37	3.21	25.83
All	10.02	16.27	38.22	25.27	10.22	100.00
Panel B: By Rating & Data Source						
Datastream	0.13	0.82	3.33	4.57	1.66	10.50
LBFI	1.44	5.99	9.11	5.01	1.43	22.99
NAIC	7.35	3.32	11.84	4.65	1.39	28.55
TRACE	1.10	6.14	13.93	11.04	5.75	37.96
All	10.02	16.27	38.22	25.27	10.22	100.00

Table B.3: Robust Check using Zero as Benchmark:
Predicting Individual Corporate Bond Excess Returns

The left panel of this table presents the time series average $\bar{R}^2_{OOS}$, and the corresponding Fama-Macbeth t-statistics as defined in equation 7. On the right panel, we report the out-of-sample R^2_{OOS} (%) of individual corporate bond excess returns as defined in equation 6. The out-of-sample R^2_{OOS} benchmark (denominator) is zero for robustness check. Panel A reports the results for investment grade, non-investment grade, and the overall bond sample. Panel B reports the results for each sub-sample of corporate bonds sorted by credit rating. Panel C reports the results for each sub-sample of corporate bonds sorted by duration length. We list the performance for eight predictive methods. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$\bar{R}^2_{OOS}$						R^2_{OOS}				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(8)	(9)	(10)
Panel A: IG/NIG sub-sample											
mean	1.2***	1.0***	1.2***				0.6	0.3	0.5		
median	0.8***	0.7***	0.8***				0.2	0.1	0.1		
lasso	3.7***	2.1***	3.3***				4.0	2.5	3.4		
ridge	3.7***	2.0***	3.4***				4.1	1.6	3.1		
pca	1.4**	1.3**	1.6**				1.0	0.1	0.6		
pls	2.8***	1.0	2.5**				3.1	1.3	2.4		
btree	5.1***	0.9	3.9***				4.9	0.5	3.1		
rf	4.9***	2.1**	4.2***				3.9	1.4	2.8		
Panel B: Credit Rating sub-sample											
	AAA	AA	A	BBB	Junk		AAA	AA	A	BBB	Junk
mean	0.4	1.0***	1.1***	1.5***	1.0***		0.5	0.5	0.6	0.8	0.3
median	0.0	0.6*	0.7***	1.1***	0.7***		-0.1	-0.1	0.1	0.4	0.1
lasso	1.7	1.2	3.0***	3.9***	2.1***		3.0	3.3	4.2	4.3	2.5
ridge	2.3**	1.0	3.1***	3.9***	2.0***		4.3	3.2	4.3	3.9	1.6
pca	0.7	-0.4	0.5	1.5**	1.3*		3.1	0.4	0.7	0.6	0.1
pls	-0.2	-1.8	2.0*	3.3***	1.0		1.3	0.5	3.9	3.8	1.3
btree	6.8***	2.3	4.3***	4.4***	0.9		8.1	4.1	4.8	4.0	0.5
rf	5.6***	3.9***	4.4***	4.3***	2.1**		6.8	3.8	3.7	3.0	1.4
Panel C: Duration sub-sample											
	Dur1	Dur2	Dur3	Dur4	Dur5		Dur1	Dur2	Dur3	Dur4	Dur5
mean	1.2***	1.2***	1.2***	1.2***	1.2***		0.3	0.4	0.6	0.5	0.5
median	0.8***	0.8***	0.8***	0.9***	0.9***		0.0	0.0	0.1	0.2	0.2
lasso	-0.1	1.8*	3.4***	3.8***	3.7***		2.7	3.0	3.9	3.6	3.4
ridge	-0.3	1.7	3.4***	3.9***	3.8***		1.8	2.6	3.8	3.4	3.1
pca	-0.1	1.0	1.7***	1.6***	1.8***		-0.4	0.3	1.0	0.8	0.8
pls	-5.1	-0.9	2.3**	3.4***	3.8***		-0.7	1.3	3.3	2.8	3.1
btree	1.0	2.8**	3.9***	3.7***	4.0***		3.0	4.1	4.3	2.6	2.4
rf	4.5***	4.3***	4.1***	4.0***	4.3***		2.4	2.8	3.4	2.7	2.8

Table B.4: Robust Check using Zero as Benchmark:
Predicting Individual Corporate Bond Excess Returns
sub-period Analysis

The left panel of this table presents the time series average $\bar{R}^2_{OOS}$, and the corresponding Fama-Macbeth t-statistics as defined in equation 7. On the right panel, we report the out-of-sample R^2_{OOS} (%) of individual corporate bond excess returns as defined in equation 6. The out-of-sample R^2_{OOS} benchmark (denominator) is zero for robustness check. The results for the overall sample are reported in table 1. This table further reports the results for the sub-periods 1998-2007 (pre-crisis), 2008-2009 (financial crisis), and 2010-2017 (post-crisis). We list the performance for eight predictive methods. For t-statistics *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$\bar{R}^2_{OOS}$			R^2_{OOS}		
	IG	NIG	All	IG	NIG	All
Panel A: 1998 - 2007						
mean	0.30	0.28***	0.38**	0.18	0.05	0.13
median	-0.06	0.16**	0.07	-0.14	-0.04	-0.10
lasso	2.57***	1.25**	2.44***	2.44	1.51	2.07
ridge	3.01***	1.14**	2.82***	2.87	1.07	2.14
pca	0.02	-0.02	0.46	-0.30	-0.65	-0.44
pls	2.45**	0.32	2.29**	2.71	1.27	2.13
btree	3.63***	-0.45	2.38**	3.76	0.16	2.30
rf	3.44***	0.50	2.70***	3.30	0.96	2.35
Panel B: 2008 - 2009						
mean	0.47***	0.20*	0.33***	0.59	0.22	0.41
median	-0.07	-0.03	-0.05	-0.05	-0.04	-0.04
lasso	4.86***	2.88**	3.77***	5.64	3.29	4.52
ridge	4.19*	1.50	2.71	5.27	1.48	3.46
pca	0.96	0.37	0.74	1.78	0.38	1.11
pls	2.02	0.80	1.15	2.79	0.74	1.81
btree	6.25***	2.44	4.38***	5.02	0.30	2.76
rf	3.92***	1.91	2.93***	2.43	0.96	1.73
Panel C: 2010 - 2017						
mean	2.52***	2.01***	2.40***	2.07	1.76	1.99
median	2.08***	1.59***	1.97***	1.52	1.24	1.44
lasso	4.80***	3.07	4.37***	5.80	4.69	5.49
ridge	4.40***	3.30*	4.18***	5.70	5.11	5.54
pca	3.12***	3.19**	3.20***	3.47	3.21	3.40
pls	3.34*	1.81	3.08**	4.99	4.19	4.77
btree	6.75***	2.32	5.76***	8.12	3.28	6.78
rf	6.94***	4.12**	6.50***	8.10	5.18	7.29