Predicting Individual Corporate Bond Returns*

Xin He **Guanhao Feng** **Junbo Wang** **Chunchi Wu**

First version: Dec. 2020; This version: Jan. 2023

Abstract

This paper documents substantial evidence of return predictability and investment gains for individual corporate bonds via machine learning. The forecast-implied long-short and market-timing strategies deliver significant risk-adjusted returns over transaction costs. Random Forest has the best performance as the ensemble of regression trees helps reduce overfitting. Using a long-span sample from 1976 to 2020, we can evaluate return predictability over business cycles and find aggregate predictors (e.g., corporate bond market return and TERM factor) that show higher forecasting power than bond characteristics. Finally, we find return predictability differs between bonds issued by private and public firms, with higher investment gains in private bonds.

JEL classification: C55, C58, G0, G1, G17.

Keywords: Bond Characteristics; Machine Learning; Aggregate Predictors; Return Predictability; Private Bonds.

----

*We are grateful to Jingzhi Huang, Injun Hwang, Guohao Tang, and Tengfei Zhang, and the conference participants at the 2021 Financial Management Association Annual Meeting, 2021 International Conference of the French Finance Association, 2022 China International Risk Forum, and 2022 Australasian Finance and Banking Conference for constructive discussions and feedback.
He (E-mail: xinh@hnu.edu.cn) is at Hunan University; Feng (E-mail: gavin.feng@cityu.edu.hk) and Wang (E-mail: jwang2@cityu.edu.hk) are at City University of Hong Kong; Wu (E-mail: chunchiw@buffalo.edu) is at State University of New York at Buffalo.

1. Introduction

Empirical methods of asset pricing tests and return forecasts have considerably improved recently (see, among others, Kelly, Pruitt, and Su, 2019; Feng, Giglio, and Xiu, 2020; Bianchi, Büchner, and Tamoni, 2021; Nagel, 2021; Kelly, Palhares, and Pruitt, 2021).[1] However, studies using new methodologies focus on the equity market, and the literature is scant for corporate bonds.[2] This is somewhat surprising as the corporate lending market is comparable in size to the equity market and is the primary source of public financing.[3] One reason could be the lack of comprehensive corporate bond transaction data before the establishment of TRACE (Trade Reporting and Compliance Engine), which commences in July 2002. Another possible reason is that a large proportion of corporate bonds are issued by private firms, whose information is not as easily accessible as publicly listed firms. Consequently, the risk-return trade-off and return predictability are much less known for private bonds.

This paper investigates individual corporate bond return predictability and evaluates the performance of various types of predictors using machine learning models. By merging multiple datasets, we construct a long-span sample of corporate bond returns from 1976 to 2020, which enables us to evaluate the return predictability over business cycles and the long-term predictor performance. We find substantial out-of-sample (OOS) return predictability in corporate bonds, especially for the recent 25 years, which can be translated into significant investment gains over transaction costs. We consider a long list of aggregate predictors proposed in the literature and find some (e.g., corporate bond market return, TERM factor, and GDP growth rate) are more useful than bond characteristics for predicting corporate bond returns.

[1] See Giglio, Kelly, and Xiu (2022) for a comprehensive survey.
[2] Notable exceptions are Lin, Wu, and Zhou (2018) and Kelly, Palhares, and Pruitt (2021).
[3] According to SIFMA capital market statistics, for the U.S. corporate bond market in 2020, the outstanding value is \$9.8 trillion, and the issuance value is \$2.3 trillion, while the corresponding values for the U.S. equity market are \$50.8 trillion and \$390.0 billion, respectively. See also Lin, Wu, and Zhou (2018).

The return predictability of individual corporate bonds has higher statistical and economic significance than those of individual stocks. The OOS R^2 is 3.50% over the testing period from 1996 to 2020, with an annualized Sharpe ratio of 3.41 for a monthly rebalanced forecast-implied investment strategy. The predictability is robust before and after the 2008-2009 financial crisis and permeates virtually all rating and maturity categories. When comparing bonds issued by public and private firms, we find no significant differences in return predictability. But the forecast-implied long-short strategy profits are larger for private bonds than public bonds, and the predictor importance differs between these two types of bonds.

The literature on corporate bond return predictability has suggested multiple aggregate predictors as time-series signals (e.g., GDP growth, inflation, and term spread) and bond characteristics as cross-sectional signals (e.g., rating, maturity, and downside risk). Lin, Wang, and Wu (2014) and Lin, Wu, and Zhou (2018) find evidence of corporate bond return predictability using forecast combination methods to evaluate a large number of predictors. However, individual bond observations are highly imbalanced because many bonds have a short history due to maturity or liquidity reasons. The lack of data and the long list of predictors make it harder to implement heterogeneous predictive modeling in individual corporate bond return forecasts.

To overcome these challenges, we employ machine learning models to consider pooled modeling and utilize a large panel.[4] We adopt a homogeneous conditional expectation functional form that applies to all bonds and use the idiosyncratic predictors to generate heterogeneous forecasts. Machine learning (ML) methods (e.g., lasso, principal component analysis, and random forest) are well known for their flexibility in dealing with high dimensionality of predictors and exploring potentially nonlinear and interactive predictor structures. However, the complexity in

[4] Recent studies on machine learning in asset pricing include Freyberger, Neuhierl, and Weber (2020) and Gu, Kelly, and Xiu (2020) for equities and Bianchi, Büchner, and Tamoni (2021), Fan et al. (2022), Huang and Shi (2022), and Huang et al. (2022) for Treasury bonds.

ML algorithms makes it difficult to conduct standard hypothesis tests and causal inferences. Based on the large panel data, this paper proposes a Fama-MacBeth (1973) type of tests to evaluate the performance of different predictive models for individual corporate bonds.

We find that machine learning models, such as random forest and lasso, outperform the well-known model combination and PLS forecasts (Lin, Wang, and Wu, 2014; Lin, Wu, and Zhou, 2018). Our study further considers aggregate predictors and distinguish them from the cross-sectional bond characteristics. The long-span sample enables us to reliably evaluate the usefulness of aggregate predictors for improving return predictions. Moreover, the extant literature on forecasting individual corporate bond returns focuses on public bonds (Ronen and Zhou, 2013; Chordia et al., 2017; Choi and Kim, 2018; Bali et al., 2022; Li et al., 2022). Our paper extends the literature by providing a comprehensive evaluation of return predictability for both public and private bonds and finds evidence of higher investment gains in private bonds.

Our paper documents several new findings that contribute to the literature. First, all machine learning models we consider in this study agree with a small subset of important predictors. The aggregate predictors, such as lagged corporate bond market return, TERM factor, and GDP growth rate, help forecast bond returns. Second, the forecast-implied long-short strategy delivers 1.86% average monthly return and a 1.83% alpha, which are economically significant. The investment returns cannot be explained by the factor models of Fama and French (1993), Bai, Bali, and Wen (2019) and Bekaert and De Santis (2021). Furthermore, we provide an alternative market-timing strategy, which generates an alpha of 1.48% that complements the long-short strategy. The monthly turnover ratios are between 30% and 150% for the trading strategies, and investment returns are larger than transaction costs.

Third, we find different predictive patterns in public and private bonds. Public bond predictability is more sensitive to aggregate predictors than bond characteristics. Stock market return and volatility, T-Bill rate, and industrial production growth rate are useful predictors of public bond returns. By contrast, term and value factors and consumption growth rate are more important predictors for private bonds. Including private bonds in the portfolio increases the investment gain, with about 0.8% higher monthly returns, and doubles the Sharpe ratio.[5] Excluding private bonds incurs a decrease in monthly investment gains as large as 0.67% from 1996 to 2020.

Fourth, we find aggregate predictors are more important than bond characteristics for public bonds, junk bonds, and long-term bonds in the predictive R-squared measures. This finding is consistent with Giesecke et al. (2011), Giesecke et al. (2014), Lin, Wang, and Wu (2014), and Lin, Wu, and Zhou (2018), who show aggregate predictors contain rich information for predicting corporate bond risk and returns. For investment impacts, the random forest long-short strategy earns 1.86% monthly returns and a 3.41 Sharpe ratio, but ignoring the aggregate predictors lowers it to 1.37% monthly returns and a 2.23 Sharpe ratio.

This paper contributes to the literature on corporate bond return predictability. Early empirical studies focus on market-timing evidence for stock and corporate bond portfolio returns (see Keim and Stambaugh, 1986 and Fama and French, 1989). Besides return predictability, the market-wide corporate default rates can be predicted by financial and aggregate predictors (see Giesecke et al. (2011)). Hong, Lin, and Wu (2012) introduce nonlinear time-series models, Lin, Wang, and Wu (2014) adopt combination forecasts, and Lin, Wu, and Zhou (2018) develop an iterated combination approach. Unlike these studies, we employ machine learning methods to predict corporate bonds using both cross-sectional and time-series predictors.

Our paper is related to Guo et al. (2022) and Bali et al. (2022). Guo et al. (2022) find past

[5] This finding is consistent with Jostova et al. (2013), who find private bonds earn higher expected returns than public bonds in momentum strategies.

yields across horizons contain predictive information on the cross-section of corporate bond returns and combine the signals with a penalized linear model. They focus on the cross-sectional signals for prediction, ignoring time-series aggregate predictors. Bali et al. (2022) apply machine learning methods with bond and stock characteristics to predict corporate bond returns via reduced-form and structural-form modeling and focus on public bonds with equity information. However, as more than 60% of bonds are issued by private firms, ignoring private bonds leads to a much smaller sample size and large information loss. Adding private bonds substantially increases the annualized Sharpe ratio from 1.83 to 3.41. The trade-off of this paper is that we do not include any equity characteristics, given the large proportion of private bonds that have no equity predictors. Unlike Guo et al. (2022) and Bali et al. (2022), who focus on the cross-sectional signals, we leverage both aggregate predictors and cross-sectional bond characteristics in return prediction.

We contribute to the bond literature by studying the difference between public and private bonds. Guo et al. (2022) find greater data transparency generates a stronger yield trend premium, and this finding holds for public bonds because public firms are more transparent. By contrast, private companies have less information accessible to investors, so their bonds are deemed riskier than public bonds. Cai, Helwege, and Warga (2007) find private bonds are more likely to be underpriced in IPO and Jostova et al. (2013) find private bonds earn higher expected returns than public bonds in the momentum strategy. Consistent with these studies, we find higher investment gains in private bonds than in public bonds using the machine learning forecast-implied strategies.

This paper is also related to studies of risk factors and anomalies for corporate bonds. Fama and French (1993) add the DEF and TERM factors to their three factors (MKT, SMB, and HML) for both stock and bond markets. Bai, Bali, and Wen (2019) introduce a four-factor model by adding downside, credit, and liquidity risk to the corporate bond market factor. Bekaert and De

Santis (2021) find the CAPM with a broad market portfolio, including treasuries and corporate bonds, is an excellent factor model for bond portfolios in cross-asset and international settings. In addition, several papers investigate various risk factors and anomalies, e.g., DEF and TERM factor betas from Gebhardt et al. (2005), liquidity from Lin, Wang, and Wu (2011) and Bao, Pan, and Wang (2011), momentum from Jostova et al. (2013), and volatility from Chung, Wang, and Wu (2019). Some researchers link the equity (Chordia et al., 2017; Choi and Kim, 2018) and option (Cao et al. (2022)) information to corporate bond returns.[6] Following the previous studies, this paper investigates corporate bond return predictability using machine learning methods with a large number of predictors and long-span data.

Our work contributes to a rapidly developing literature on empirical asset pricing via machine learning. Recent papers find stock return predictable using machine learning and deep learning methods, e.g., Freyberger, Neuhierl, and Weber (2020), Gu, Kelly, and Xiu (2020, GKX2020 henceforth), Avramov, Cheng, and Metzker (2022), among others. GKX2020 and Li and Rossi (2020) find boosted regression tree and random forest models are strong machine learning models for predicting stock and mutual fund returns. Bianchi, Büchner, and Tamoni (2021), Fan et al. (2022), Huang and Shi (2022), and Huang et al. (2022) find improved predictability on treasury bond returns using aggregate or macroeconomic information. Kelly, Pruitt, and Su (2019) and Feng, Giglio, and Xiu (2020) develop different machine learning methods for cross-sectional asset pricing tests. In the corporate bond literature, Kelly, Palhares, and Pruitt (2021) and Feng et al. (2022) create latent corporate bond factor models via PCA and deep learning, Li et al. (2022) apply machine learning to corporate bond returns in international markets, and He et al. (2022) propose a combination approach on corporate bond basis portfolios and individual bond returns.

[6] See Huang and Shi (2021) for a survey.

The rest of the paper is organized as follows. Section 2 describes the data and predictive model designs. Section 3 provides empirical evidence of return predictability. Section 4 evaluates the variable importance of predictors. Section 5 implements the investment strategies based on the forecasts. Section 6 summarizes our main findings and concludes the paper.

2. Methodology

2.1 Predictive model design

Following GKX2020, we characterize excess returns of assets by a model with an additive prediction error:

$$r_{i,t+1} = E\left(r_{i,t+1}\right) + \epsilon_{i,t+1}, \tag{1}$$

where bonds are indexed as $i = 1, \cdots, N_t$ and months by $t = 1, \cdots, T$. We use a panel approach to predict bond returns, assuming a homogeneous conditional expectation functional form that applies to all bond-return observations. Indeed, there is a trade-off between predictability heterogeneity and estimation accuracy. We do not adopt time-series modeling for each bond because individual bond returns are highly imbalanced. Many bonds have a short historical sample for maturity or liquidity reasons. The panel approach is a common practice for modeling individual asset returns utilizing a large dimension of bond or firm characteristics (see GKX2020; Leippold, Wang, and Zhou, 2022). Therefore, any discovered return predictability is mainly from the common predictors. The conditional expectation can be expressed as

$$E\left(r_{i,t+1}\right) = g_t\left(z_{i,t}, x_t\right), \tag{2}$$

where bond characteristics are denoted as $z_{i,t}$ and aggregate predictors as x_t. At the end of each year, we recursively estimate the model $g_t(\cdot)$ with a rolling window of the past 20 years. Then, we fix the model for the next 12 months for prediction. Although the model is updated annually, the predictors are updated monthly.[7]

We survey eight predictive models of four groups, including combination forecast (mean and median combination), penalized linear regression (lasso and ridge), dimension reduction (PCA and PLS), and ensemble tree (boosted tree and random forest) models. For brevity, we only report mean combination, lasso, PLS, and random forest for some results, representing four different types of predictive models. We report more details about these models and parameter tuning in Appendix A. The training and predicting schemes follow GKX2020 without any look-ahead bias.

2.2 Performance evaluation metrics

2.2.1 Out-of-sample R^2

The predictive performance is evaluated by the out-of-sample R^2:

$$R^2_{OOS} = 1 - \frac{\sum_{i,t}\left(r_{i,t} - \hat{r}_{i,t}\right)^2}{\sum_{i,t}\left(r_{i,t} - \bar{r}_{i,t}\right)^2}, \tag{3}$$

where *t* denotes a period in the hold-out sample. This R^2_{OOS} sums prediction errors across all assets and periods to assess the reduction in mean squared forecast errors relative to a benchmark forecast. Unlike GKX2020, who use a naive zero forecast as the benchmark in the denominator, we use the 20-year moving average excess return of the corresponding rating sorted portfolio in the denominator.[8] The rating sorted portfolio moving average return we use is a stronger benchmark than the naive zero forecast and the factor model forecasts.[9]

---

[7] We align all the predictors in monthly frequency, although some are updated quarterly, e.g., consumption growth rate and GDP growth rate. We don't update the model monthly, because it takes 12 times more computation than annual updating, but offers marginal revision of models.

[8] We sort the bonds into five rating portfolios, AAA, AA, A, BBB, and Junk. This is a stronger benchmark than the moving average of the bond market excess return and zero (used in GKX2020 for stock prediction). For example, the prediction benchmark for an AAA bond is the 20-year moving average excess return of the AAA portfolio.

[9] In the last two rows of Tables 2, 3, and 4, we report the out-of-sample R^2 of naive zero forecasts, BS3 (Bekaert and De Santis, 2021), FF3 (Fama and French, 1993), and FF5 (Fama and French, 1993); as a result, we find all the numbers are negative or marginally positive. So, the naive zero forecast, BS3, FF3, and FF5 forecast are weaker than the rating sorted portfolio benchmark. BS3 include the equity market factor, corporate bond market factor, and treasury bond market factor, see footnote 1 of Bekaert and De Santis (2021); FF3 include MKT, DEF, and TERM; FF5 include MKT, SMB, HML, DEF, and TERM.

2.2.2 *Fama-MacBeth out-of-sample R^2*

The Fama and MacBeth (1973) method is standard in the empirical asset pricing literature. We exploit the advantage of this method in increasing test efficiency by providing a Fama-MacBeth-type out-of-sample R^2_{OOS} for model evaluation. Specifically, for each month *t*, we calculate

$$R^2_{OOS,t} = 1 - \frac{\sum_i (r_{i,t} - \hat{r}_{i,t})^2}{\sum_i (r_{i,t} - \bar{r}_{i,t})^2}, \tag{4}$$

and obtain a time series of $\{R^2_{OOS,t}\}_{t=1}^{T}$.[10] The Fama-MacBeth design is feasible because we have big panel data, which are long in time series and large in the cross-section. The Fama-MacBeth average out-of-sample $\bar{R}^2_{OOS}$ is

$$\bar{R}^2_{OOS} = \frac{1}{T} \sum_{t=1}^{T} R^2_{OOS,t}\,. \tag{5}$$

Unlike R^2_{OOS}, the Fama-MacBeth modeling allows us to evaluate the statistical significance of R^2_{OOS} with Newey-West (1987) robust standard errors. The testing hypotheses are

$$H_0 \colon \bar{R}^2_{OOS} = 0 \; and \; H_1 \colon \bar{R}^2_{OOS} \neq 0.$$

Evaluating the forecast performance difference for different sets of results is useful, e.g., $\{R^2_{OOS,t;1}\}_{t=1}^{T}$ and $\{R^2_{OOS,t;2}\}_{t=1}^{T}$ for two subsamples. The null and alternative hypotheses are

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

We calculate the *t*-statistic for the average of the time-series difference:

$$R^2_{OOS:\Delta} = R^2_{OOS:1} - R^2_{OOS:2}. \tag{6}$$

and then conduct hypothesis testing.

---

[10] Chen, Roussanov, and Wang (2021), Li et al. (2022), and Kelly, Palhares, and Pruitt (2021) use similar applications of Fama-MacBeth-type R^2 measures, that first aggregate information in the cross-section and then take average in the time series.

2.2.3 *Variable importance*

We identify important predictors with high predictive power for individual corporate bond returns. Like GKX2020, we construct a new predictor dataset by setting all values of predictor *j* to 0 and keeping the remaining predictors unchanged. Next, we give a new prediction $\tilde{r}_{i,t}^{[j]}$ with the initially fitted model and the new predictor dataset. The variable importance is measured by the reduction in R^2_{OOS},

$$R_{OOS}^{2,[j]} = 1 - \frac{\sum_{i,t}\left(r_{i,t} - \tilde{r}_{i,t}^{[j]}\right)^2}{\sum_{i,t}\left(r_{i,t} - \bar{r}_{i,t}\right)^2}, \tag{7}$$

$$vi^{[j]} = R^2_{OOS} - R^{2,[j]}_{OOS}. \tag{8}$$

Our variable importance measure is different from GKX2020 and more suitable for our empirical purposes. The variable importance in GKX2020 measures the contribution of each predictor to the in-sample model fitting, and they report the relative variable importance, which is the ranking among all predictors. However, one has no idea about the real impact of predictors on out-of-sample prediction with the relative in-sample variable importance of GKX2020. By contrast, our out-of-sample variable importance is about the out-of-sample contribution of each predictor. We provide both the raw measure and relative measure in empirical study. The advantage of the out-of-sample variable importance is that we can quantify the real impact of each predictor on the out-of-sample prediction. The variable importance would be positive if a predictor benefits out-of-sample prediction conditional on the other variables, and vice versa.

3. Evidence on return predictability

3.1 *Data*

3.1.1 *Corporate bond returns*

The corporate bond returns data are from four sources: the Lehman Brothers Fixed Income

(LBFI) database, DataStream, the National Association of Insurance Commissioners (NAIC) database, and the Trade Reporting and Compliance Engine (TRACE) database. We consolidate the data from these sources to get a large corporate bond return sample from 1973 to 2020. We follow a priority ranking and keep only one observation when duplicates arise from multiple sources. The high to low priority ranking is TRACE, NAIC, LBFI, and DataStream, following Lin, Wang, and Wu (2014) and Lin, Wu, and Zhou (2018). To avoid confounding effects, we focus on straight bonds in empirical tests. That is, we exclude any asset-backed bond observations, any option-embedded, non-U.S.-dollar-denominated bonds, bonds with unusual coupons, and bonds with a maturity of less than one year or over 30 years. The monthly corporate bond return at time t is calculated as follows:

$$R_t = \frac{(P_t + A_t) + C_t - (P_{t-1} + A_{t-1})}{P_{t-1} + A_{t-1}}, \tag{9}$$

where P_t is the price, A_t is the accrued interest, and C_t is the coupon payment at time t. We obtain excess returns by subtracting the three-month T-Bill rate from the raw return. Table 1 provides summary statistics of the sample.

[Insert Table 1 here]

3.1.2 Bond predictors

We collect a comprehensive set of predictors, including 20 aggregate predictors and 20 corporate bond characteristics, with detailed descriptions in Table B.1 of Appendix B.[11] Welch and Goyal (2008) identify a list of aggregate predictors for equity premium prediction. Lin, Wang, and Wu (2014) and Lin, Wu, and Zhou (2018) find that aggregate predictors contain rich information for predicting corporate bond returns. Giesecke et al. (2011, 2014) and Azizpour et al.

[11] We do not include any equity characteristics, because they are only applicable to the corporate bonds issued by publicly listed firms.

(2018) find that aggregate predictors can predict corporate default rates of the corporate bond market. Our aggregate predictors cover three main categories: economic indicators (e.g., GDP growth, inflation), bond market variables (e.g., term spread, forward factor), and equity market variables (e.g., S&P 500 Index Returns, S&P 500 Index Earnings-to-Price Ratio).

The corporate bond characteristic sample covers three categories: fundamental characteristics (e.g., rating, maturity), return-distribution characteristics (e.g., momentum, downside risk), and covariances with common risk factors (e.g., regression betas on DEF and TERM factors). The return sample period is from January 1973 through September 2020. The final sample of predictors starts from January 1976, because we need three years of data to calculate the return-based characteristics. In empirical analysis, the OOS evaluation period is 25 years, from January 1996 to September 2020.

3.2 *Out-of-sample Predictability*

We find substantial evidence of OOS return predictability over the 25-year out-of-sample test period from 1996 to 2020. Table 2 reports the predictability results. The first three columns report the monthly average $\bar{R}^2_{OOS}$ (see (5)) and the *t*-test significance for the return predictability, and the last three columns report the summary R^2_{OOS} (see (3)). Based on the R^2_{OOS} measure, corporate bond returns are predictable by all surveyed models except the median combination. Further, $\bar{R}^2_{OOS}$ shows positive numbers for all models except the median combination. The null hypothesis of no predictability is rejected at the 1% level for mean combination, lasso, ridge, PLS, boosted tree, and random forest. Random forest, which is nonlinear, is the best-performing model with 3.94% $\bar{R}^2_{OOS}$ and 3.50% R^2_{OOS}. Lasso is the best among the linear models, with 3.06% $\bar{R}^2_{OOS}$ and 3.35% R^2_{OOS}. In addition, we investigate the predictability of IG (investment-grade) and NIG (non-investment-grade) bonds, separately. We find the IG bonds are more predictable than NIG bonds for many predictive models. Overall, we find strong predictive evidence of corporate bond returns for the whole sample.

The predictability measure is relative to the benchmark prediction in the denominators of (3) and (5). Admitting that many benchmark predictions exist in the corporate bond market, we list four and compare them with our rating sorted portfolio benchmark in the last four rows of Table 2. The four benchmark predictions are naïve zero prediction, three factor model implied return predictions[12] for BS3 (equity market factor, corporate bond market factor, and treasury bond market factor) in Bekaert and De Santis (2021), FF3 (MKT, DEF, and TERM) and FF5 (MKT, SMB, HML, DEF, and TERM) in Fama and French (1993). As a result, we find none of the four benchmark predictions significantly outperforms our rating sorted portfolio benchmark, as the values of $\bar{R}^2_{OOS}$ are either close to zero or significantly negative.

[Insert Table 2 here]

3.3 Differential predictability in time series and cross-section

To investigate the potential heterogeneity in model predictability, we conduct subsample analysis for different rating and maturity groups in the cross-section. Also, we assess the robustness of return predictability before and after the 2008-2009 financial crisis. The OOS (out-of-sample) period from 1996 to 2020 is divided into three sub-periods: pre-crisis (1996-2007), crisis (2008-2009), and post-crisis (2010-2020).

Table 3 reports the predictability of five rating groups for different periods. Panel A shows that all five rating groups are predictable in terms of R^2_{OOS} for mean combination, lasso, ridge, and random forest. Lasso and random forest are better candidates for predictive models, delivering significantly positive $\bar{R}^2_{OOS}$ for all rating groups. Again, according to the last four rows of Table 3

[12] The factor model predictions are the product of bond beta, based on past 60-month rolling-window beta estimation, and factor risk price estimated as the 20-year rolling-window average return of the factors.

Panel A, none of the four benchmark predictions significantly outperforms our rating sorted portfolio benchmark in any rating groups.

The results for the pre-crisis period are reported in Panel B of Table 3, which shows large values of return predictability. The 2008 global financial crisis dramatically impacts predictability, especially for high-rating bonds. In Panels C (crisis period) and D (post-crisis period), lasso and random forest have negative or insignificantly positive $\bar{R}^2_{OOS}$ for the AAA group. For A, BBB, and NIG groups, the predictability is robust over the crisis period.

Table 4 reports the results for different maturity groups. We sort the bonds into five groups by their maturities in the cross-section. Each group contains the same number of bonds each month. TMT1 is the short-maturity group, and TMT5 is the long-maturity group. The results show that random forest is the best model for the OOS period of 1996 to 2020 and produces significantly positive results for all five maturity groups. Lasso and ridge produce positive numbers for the five maturity groups in terms of R^2_{OOS}. However, their predictability is inconsistent over time. By contrast, boosted tree and random forest handle the heterogeneous characteristics of different maturities effectively and predict all groups well. Still, none of the four benchmark predictions significantly outperforms our rating sorted portfolio benchmark in any maturity groups, according to the last four rows of Table 4 Panel A.

The impact of the 2008 global financial crisis is not manifested in maturity groups. In Panels C and D, we report for the periods during and after 2008 and find the predictability evidence is robust for the random forest. However, the performance of lasso and PLS deteriorates for short-maturity bonds. It seems nonlinear models are more reliable over the 2008 financial crisis period than linear models.

[Insert Tables 3 and 4 here]

To summarize, junk bonds are more challenging to predict than investment-grade bonds. The 2008 global financial crisis greatly impacts the AAA bond predictability. The maturity groups are almost equally predictable if we use the random forest, and are not affected by the 2008 global financial crisis. Overall, the random forest is a robust model for the 2008 financial crisis and the normal period.

3.4 Predictability on private and public bonds

Chordia et al. (2017) and Choi and Kim (2018) find corporate bond returns are predictable using equity characteristics. By contrast, Bali et al. (2022) find the equity characteristics redundant, controlling for bond characteristics in the reduced-form predictive modeling. These two papers obtain different empirical conclusions but share one thing: they both ignore the private-bond observations or impute equity information of private bonds with the cross-sectional median. In this paper, we consider all corporate bonds, either issued by public or private firms. If we include equity information to predict bond returns, we will face a massive missing data problem for private bonds. Therefore, we exclude equity information and only include bond characteristics and aggregate predictors that are available for all corporate bonds.

We investigate the difference in the predictability of private and public bonds. Private and public company bonds would be equally predictable if the financial market is efficient. If equity characteristics provide additional signals beyond bond characteristics to return predictability, dropping these predictors will decrease the predictability of public company bond returns. Thus, the predictability of public bonds could be smaller than private bonds, when excluding equity characteristics. Given the previous predictive modeling setup, we test the following hypothesis:

$$H0\colon \bar{R}^2_{OOS;Private} = \bar{R}^2_{OOS;Public} \; and \; H1\colon \bar{R}^2_{OOS;Private} \neq \bar{R}^2_{OOS;Public}.$$

Table 5 reports the predictability for public and private company bonds separately with the same training model. In the third column, we calculate the average of the differences in the time series between the $R^2_{OOS,t}$ of public and private bonds with the Fama-MacBeth *t*-statistics, (see (6)). Panel A shows private company bonds have larger R^2_{OOS} than public company bonds for lasso, PLS, and random forest, and smaller $\bar{R}^2_{OOS}$ for mean combination. However, all these positive and negative differences are statistically insignificant from zero.

Panels B, C, and D report for time subsamples of before, during, and after the 2008 global financial crisis. Private bonds usually have a larger $\bar{R}^2_{OOS}$, and thus are more predictable than public bonds. However, the *t*-tests indicate the differences are statistically insignificant. Panels E and F report results for rating subsamples. The differences between private and public bond predictability are statistically insignificant for IG and NIG bonds, except that public bonds are more predictable for the PLS model in the NIG subsample at the 10% confidence level. Panels G and H report the results for maturity subsamples. We find all the numbers are statistically insignificant from zero. Overall, we cannot reject the null hypothesis that private and public bonds have similar predictability, and the conclusion holds for different rating and maturity subsamples. Our finding is consistent with Ronen and Zhou (2013) finding that stock trading does not lead the corporate bond market when institutional trades are dominant and other bond trading features are controlled.

[Insert Table 5 here]

4. Variable importance for return predictability

We further investigate the marginal contribution for 20 aggregate predictors and 20 bond characteristics. More importantly, we are interested in their net contribution to the OOS prediction. To this end, we apply the out-of-sample variable importance measure in Section 2.2.3, which is different from the variable importance measure of GKX2020. Also, we highlight the predictive power of aggregate predictors, which align with Lin, Wang, and Wu (2014), and Lin, Wu, and

Zhou (2018) and complement Chordia et al. (2017), Choi and Kim (2018), and Bali et al. (2022), who use bond and equity characteristics for returns prediction.

4.1 Relative variable importance

Figure 1 reports the out-of-sample relative variable importance for all models. The relative variable importance within each model is normalized to sum to 1. We order the predictors by their average ranks across all models, where the important ones are at the top and the least important ones are at the bottom. The results show that machine learning methods produce a small set of aggregate predictors (e.g., corporate bond market excess returns, TERM factor, and GDP growth rate) and bond characteristics (e.g., downside risk, short-term reversal, and skewness). Our results are consistent with Giesecke et al. (2011), who show stock volatility, GDP growth, and industrial production growth help forecast corporate bond market-level default risks. For the bond characteristics group, downside risk and short-term reversal are important, consistent with Bai, Bali, and Wen (2019).

Figure 2 reports the top 10 most important variables of each model. Aggregate predictors account for nine (seven) of the top 10 predictors for random forest (lasso). Therefore, aggregate predictors appear to be more important than bond characteristics in OOS prediction.

[Insert Figures 1 and 2 here]

4.2 Variable importance

We next investigate the raw out-of-sample variable importance, which contains more information than the relative measure. Figure 3 shows each model's heat map of variable importance. A positive number is in green, a negative number is in red, and a number close to zero is in yellow. The order of variables follows Figure 1. Many predictors provide a tiny contribution to OOS prediction, and some predictors in dark red even harm OOS prediction. We find that some “relatively less important” predictors in Figure 1 are harmful for OOS prediction, for example, stock market liquidity and equity size factor SMB in aggregate predictors and rating and coupon rate in bond characteristics.

These findings are ignored by the relative variable importance measure, which is one of the most popular metrics in machine learning in the finance literature. Our finding is consistent with Welch and Goyal (2008), who show many predictors have predictive power for in-sample prediction but do not pass the OOS test. Nevertheless, we find a small group of important predictors that most models agree with, for example, corporate bond market returns, TERM factor, and GDP growth rates in aggregate predictors, and downside risk, short-term reversal, and skewness in bond characteristics.

[Insert Figure 3 here]

4.3 Predictive performance of random forest

Random forest seems to be the best model for prediction accuracy, as discussed in Section 3. An often-mentioned advantage of tree-based models is the clear graphical interpretation. However, ensemble tree models, such as random forest, do not offer such an advantage because it is hard, for example, to visualize and interpret 1,000 trees in a forest. The following discussion helps us understand the variable importance of aggregate predictors.

We refit the random forest model for two cases, one with only aggregate predictors and the other with only characteristics, and then compare their R^2_{OOS} differences. Figure 4 shows the changes of R^2_{OOS} from only using characteristics to only using aggregate predictors. The green color denotes an increase in importance, while red denotes a decrease from bond characteristics to aggregate predictors. A deep color means the magnitude of the difference is large, and a light color means the difference is small. In Panel A, the aggregate predictor group is more important than the characteristics for the overall sample and public bonds. But, the two groups of predictors are almost equally important for private bonds.

Panel B shows aggregate predictors are more important for low-rating bonds, and characteristics are more important for AAA bonds. In Panel C, aggregate predictors are more important for long-term bonds and tend to favor high-risk and high-volatility bonds, e.g., junk and long-term bonds. A plausible explanation is that junk bonds are more sensitive to market-level default risk, and long-term bonds are more sensitive to term structure changes.

With a particular interest in aggregate predictors, we pick the 10 most important aggregate predictors [13] and report their variable importance changes between two groups of assets. Figure 5 Panel A shows public bonds rely more on value factor (HML) than private bonds. We conjecture that stock value information can be more quickly transmitted to the corporate bond market because public bonds are more transparent (see Cai, Helwege, and Warga, 2007; Guo et al., 2022). We find private bonds are favored by stock market volatility, stock market returns, and the T-Bill rate, which is consistent with Giesecke et al. (2011), who show these three predictors help predict market-level default risk. We take one more step from Giesecke et al. (2011) and find this predictive relationship is stronger for private bonds.

Figure 5 Panel B reports the results for rating groups. We find stock market returns, GDP growth rate, and stock market volatility are more important for junk bonds, but the TERM factor is in the opposite direction. We conjecture that junk bonds are more sensitive to stock market variations than investment-grade bonds as junk bonds behave more like equities. Panel C shows the top 10 aggregate predictors are more important for long-term bonds than short-term bonds.

[Insert Figures 4 and 5 here]

[13] Table B.2 reports the relative variable importance of aggregate predictors.

5 Forecast-implied investment strategies

5.1 Long-short strategy

We construct model-forecast-implied long-short portfolios to assess the investment gain. We sort the individual bonds into quintiles based on the predicted returns. For each quintile, we construct a value-weighted portfolio and rebalance monthly. A long-short portfolio is to long the highest (quintile 5) portfolio and short the lowest (quintile 1) portfolio. Figure B.2 reports the cumulative portfolio returns for four predictive methods: mean combination, lasso, PLS, and random forest. We find monotonic spreads for sorted portfolio returns in the subplots, confirming the predictive evidence. The long-short portfolio return spread is substantial, and the premium mainly comes from the long side.

Table 6 reports multiple performance measures of the long-short portfolio, including the average return, annualized Sharpe ratio, risk-adjusted performance (alpha), and alpha *t*-statistics significance. The columns report the investment performance of five subsamples, including the whole bond universe, IG bonds, NIG bonds, private bonds, and public bonds. Specifically, we reform the strategies with a subsample of bonds as the investment pool, but the predicted returns are produced from the same model with the whole cross-section of bonds. We find the nonlinear methods (random forest) outperform the linear methods (lasso and PLS). Random forest achieves the best performance with 1.86% monthly average return for the overall sample, 1.80% for IG, 1.70% for NIG, 2.02% for private bonds, and 1.19% for public bonds. Moreover, random forest produces a 3.41 annualized Sharpe ratio for the whole sample, 3.73 for IG, and 3.64 for private bonds. These strategies cannot be explained by the five-factor model of Fama and French (1993) or the four-factor model of Bai, Bali, and Wen (2019).

We compare the investment gains for public and private company bonds. The average return from private bonds is higher than that of public bonds. The random forest strategy earns 0.80% higher monthly returns and doubles the Sharpe ratio on private bonds relative to public bonds. Our finding confirms Jostova et al. (2013) that the investment opportunity is bigger in the set of private bonds than that in public bonds.

[Insert Table 6 here]

5.2 Market-timing strategy

Market-timing is difficult in the equity market, according to Jiang (2003), who find little market-timing ability among actively managed U.S. equity funds. Here, we provide a market-timing strategy and show its great potential in corporate bond investments.

We can time the market by longing the bonds with positive return predictions and short the bonds with negative return predictions. Our market-timing strategy is a long-short portfolio combining the long and short legs. But the logic of the market-timing strategy differs from the long-short strategy in Section 5.1. Ideally, the market-timing strategy longs all the bonds in a market boom and shorts all the bonds during a market bust. In normal periods, the market-timing strategy takes both long and short positions. By contrast, the long-short strategy holds the fixed position for the long and short legs, for example, longing top 20% and shorting bottom 20%.

Figure B.3 shows the cumulative returns of the long-leg, short-leg, and market-timing strategies. We find the market-timing strategies of lasso and PLS have higher expected returns than the long-short strategy counterparts. Table 7 reports the investment performance. For lasso and PLS, the higher expected returns of market-timing over long-short strategy mainly come from IG bonds. The random forest strategy earns 1.51% monthly returns and a 1.72 Sharpe ratio. Moreover, the turnover ratio of the random forest strategy is 35% less for market timing than for the long-short strategy, which incurs a lower transaction cost.

We compare the investment gains of the market-timing strategies for public and private company bonds. For multiple models, private bonds provide higher expected returns than public bonds. The random forest strategy earns about 0.60% higher monthly returns and improves the Sharpe ratio on private bonds relative to public bonds.

[Insert Table 7 here]

5.3 Investing with partial predictors

Variable importance has been investigated in Section 4, regarding predictive accuracy. We are also interested in the economic gain from the predictors. Table 8 reports the investment performance if we only use aggregate predictors or only use corporate bond characteristics to train and predict bond returns. The investment universe is all bonds in the cross-section.

If we only use aggregate predictors, we are predicting the market-level returns, and no cross-sectional variation exists in the return predictions in a particular month. Thus, the long-short strategy is not feasible. But the time-series variation is still there, and we can apply the market-timing strategy. The market-timing strategy with only aggregate predictors earns 0.47% (0.39%) monthly returns and a 1.55 (1.23) annual Sharpe ratio for lasso (random forest), with a large alpha on Fama-French five factors.

With only bond characteristics, the long-short strategy of random forest earns 1.37% monthly with a 2.23 Sharpe ratio. However, adding the aggregate predictor increases the average return to 1.86% and the Sharpe ratio to 3.41. The market-timing strategy of the random forest also benefits from adding aggregate predictors, with an increase in average returns from 0.64% to 1.51%. Overall, we find aggregate predictors and bond characteristics contribute to the investment gain. Dropping either of them would incur a lower investment performance.

[Insert Table 8 here]

5.4 Transaction cost

The literature has suggested different measures of transaction cost for the corporate bond market, which is non-negligible for trading. According to Edwards, Harris, and Piwowar (2007), the transaction cost for corporate bonds ranges from 20 bps to 80 bps for different levels of trade size. Also, they find bonds with smaller issue size have higher transaction costs, conditional on the trade size. Bao, Pan, and Wang (2011) and Bali et al. (2022) use the Roll (1984) measure of effective spread to proxy for the corporate bond transaction cost. However, estimating the transaction cost for our early samples back to the 1970s is infeasible because of the lack of data. Instead, we follow Leippold, Wang, and Zhou (2022) to report the turnover ratio of our strategy and the investment performance net of transactions for different transaction cost levels.

Table 9 reports the average turnover percentages, monthly returns, and the Sharpe ratios for different levels of transaction costs ranging from 0 bps to 80 bps. The monthly rebalanced turnover ratio is low for the mean combination strategy but more than 100% for lasso and random forest. There are still sizeable investment gains net of the transaction cost. The value-weighted random forest long-short (market-timing) strategy earns 0.73% (0.67%) monthly returns with a 1.33 (0.77) annualized Sharpe ratio, even if we charge 80 bps for each transaction.

Our strategies do not rely on small bonds, which have relatively higher transaction costs (see Edwards, Harris, and Piwowar (2007)). In the appendix, we report in Figure B.4 for the equal-weighted long-short strategy and in Figure B.5 for the equal-weighted market-timing strategy. The equal-weighted strategies earn lower expected returns than their value-weighted counterparts. Thus, the investment gains arise more from the large bonds with lower transaction costs instead of the small bonds. These findings relieve the concerns about transaction costs of the value-weighted investment strategies.

[Insert Table 9 here]

6 Conclusion

This paper shows robust positive findings for adopting machine learning methods in predicting individual corporate bond returns using a comprehensive sample from 1976 to 2020. The forecasting results align with and contribute to the literature on bond return predictability and machine learning in asset pricing. Linear machine learning models produce a positive return prediction performance. However, we find nonlinear methods, such as random forest, deliver more consistent and superior performance across bond subsamples and subperiods. These promising bond return predictability results via machine learning are important to academic and practitioner researchers.

Considering bonds issued by private companies, we find their useful return predictors differ from bonds issued by public firms. The predictability of public bond returns is more sensitive to the T-Bill rate, equity market return, and volatility. In terms of investment gains, we find private bonds deliver higher returns than public bonds from investment strategies. These findings provide compelling economic evidence that private bonds are important for corporate bond investment. For 20 aggregate predictors and 20 bond characteristics considered for return prediction, we find that aggregate predictors contain substantially more information for return predictability. Excluding aggregate predictors leads to a decline in predictive accuracy and investment gains. Finally, our market-timing and long-short strategies generate sizeable investment gains net of transaction costs. The results show that machine learning enhances investment returns in the corporate bond market.

References

Avramov, D., Cheng, S., Metzker, L., 2022. Machine learning versus economic restrictions: Evidence from stock return predictability. *Management Science*, forthcoming.

Azizpour, S., Giesecke, K., Schwenkler, G., 2018. Exploring the sources of default clustering. *Journal of Financial Economics* 129, 154–183.

Bai, J., Bali, T.G., Wen, Q., 2016, Do the distributional characteristics of corporate bonds predict their future returns?. Georgetown University Unpublished working paper.

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

Bali, T.G., Goyal, A., Huang, D., Jiang, F., Wen, Q., 2022. Predicting corporate bond returns: Merton meets machine learning. Georgetown University Unpublished working paper.

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

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

Bekaert, G., De Santis, R.A., 2021. Risk and return in international corporate bond markets. *Journal of International Financial Markets, Institutions and Money* 72, 101338.

Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premiums with machine learning. *Review of Financial Studies* 34, 1046–1089.

Cai, N., Helwege, J., Warga, A., 2007. Underpricing in the corporate bond market. *Review of Financial Studies* 20, 2021–2046.

Cao, J., Goyal, A., Xiao, X., Zhan, X., 2022. Implied volatility changes and corporate bond returns. *Management Science*, forthcoming.

Chen, Q., Roussanov, N., Wang, X., 2021. Semiparametric conditional factor models: Estimation and inference. University of Pennsylvania Unpublish working paper.

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

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

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

Cochrane, J.H., Piazzesi, M., 2005. Bond Risk Premia. American Economic Review 95, 138-160.

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

Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25, 23–49.

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

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

Fan, Y., Feng, G., Fulop, A., Li, J., 2022. Real-time macro information and bond return predictability: Does deep learning help?. City University of Hong Kong Unpublished working paper.

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

Feng, G., Jiang, L., Li, J., Song, Y., 2022. Deep Tangency Portfolio. City University of Hong Kong Unpublished working paper.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. *Review of Financial Studies* 33, 2326–2377.

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

Giesecke, K., Longstaff, F.A., Schaefer, S., Strebulaev, I., 2011. Corporate bond default risk: A 150-year perspective. *Journal of Financial Economics* 102, 233–250.

Giesecke, K., Longstaff, F.A., Schaefer, S., Strebulaev, I.A., 2014. Macroeconomic effects of corporate default crisis: A long-term perspective. *Journal of Financial Economics* 111, 297–310.

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

Giglio, S., Kelly, B., and Xiu, D., 2022. Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics* 14, 337-368.

Guo, X., Lin, H., Wu, C., Zhou, G., 2022. Predictive information in corporate bond yields. *Journal of Financial Markets* 59, 100687.

He, X., Feng, G., Wang, J., Wu, C., 2022. Corporate Bond Pricing via Benchmark Combination Model. City University of Hong Kong Unpublished working paper.

Hong, Y., Lin, H., Wu, C., 2012. Are corporate bond market returns predictable? *Journal of Banking & Finance* 36, 2216–2232.

Huang, D., Jiang, F., Li, K., Tong, G., Zhou, G., 2022. Are bond returns predictable with realtime macro data? Singapore Management University Unpublished working paper.

Huang, J.Z., Shi, Z., 2021. What do we know about corporate bond returns? *Annual Review of Financial Economics* 13, 363–399.

Huang, J.Z., Shi, Z., 2022. Machine-learning-based return predictors and the spanning controversy in macro-finance. *Management Science*, forthcoming.

Jiang, W., 2003. A nonparametric test of market timing. *Journal of Empirical Finance* 10, 399–425.

Jostova, G., Nikolova, S., Philipov, A., Stahel, C.W., 2013. Momentum in corporate bond returns.

Review of Financial Studies 26, 1649–1693.

Keim, D.B., Stambaugh, R.F., 1986. Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17, 357–390.

Kelly, B.T., Palhares, D., Pruitt, S., 2021. Modeling corporate bond returns. *Journal of Finance*, forthcoming.

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

Leippold, M., Wang, Q., Zhou, W., 2022. Machine learning in the Chinese stock market. *Journal of Financial Economics* 145, 64–82.

Li, B., Rossi, A.G., 2020. Selecting mutual funds from the stocks they hold: A machine learning approach. Georgetown University Unpublished working paper.

Li, D., Lu, L., Qi, Z., Zhou, G., 2022. International corporate bond market: Uncovering risks using machine learning. Washington University in St. Louis Unpublished working paper.

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

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

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

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

Nagel, S., 2021. Machine Learning in Asset Pricing. Princeton University Press.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

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

Ronen, T., Zhou, X., 2013. Trade and information in the corporate bond market. *Journal of Financial Markets* 16, 61-103.

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

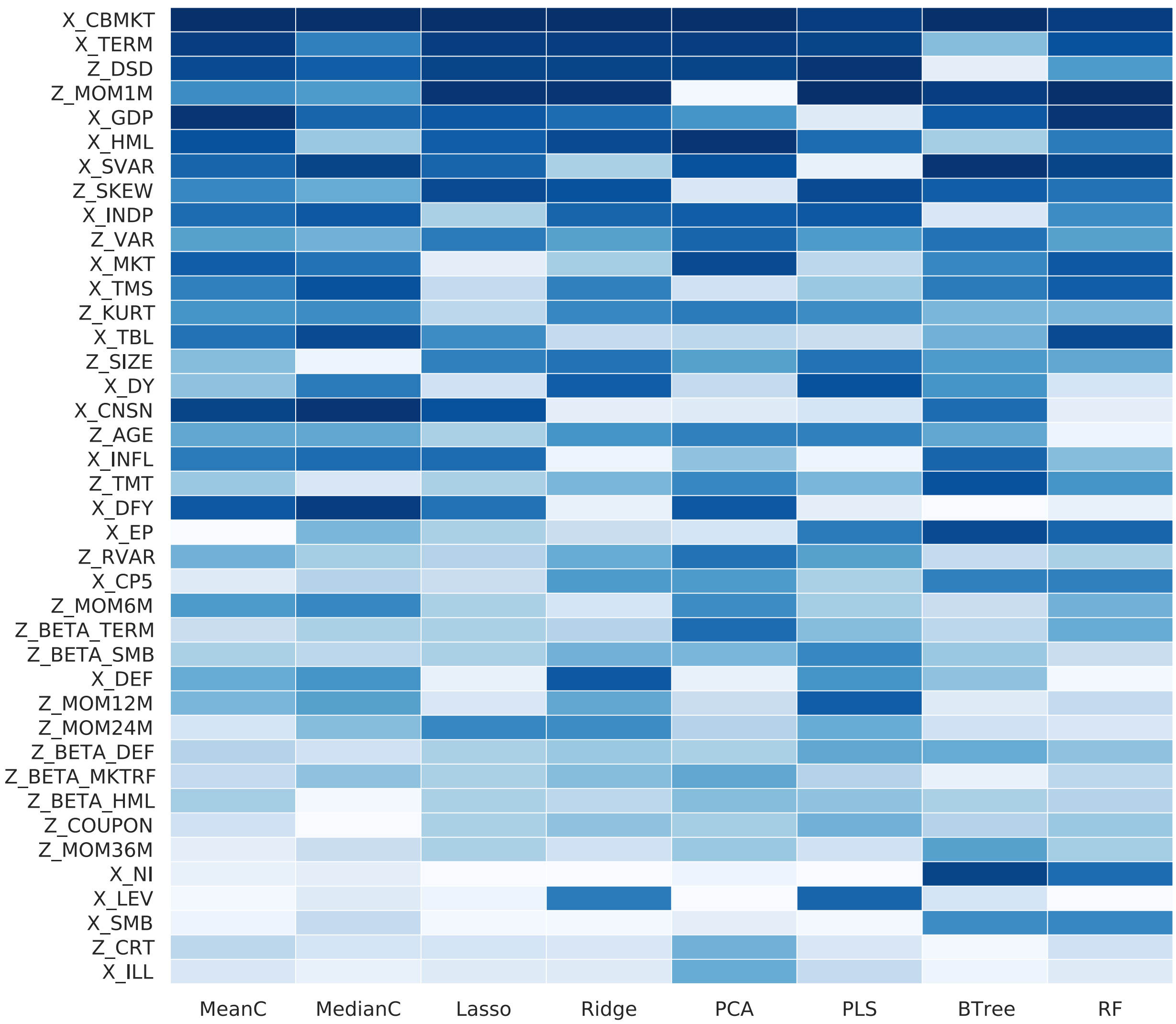

Figure 1: Heat Map of Relative Variable Importance

Forty predictors are ranked in terms of overall model contribution. We add the prefix "X" for aggregate predictors and "Z" for bond characteristics to distinguish the two groups of predictors. Predictors are ordered based on the average rank of variable importance across all models, with the most important characteristics at the top and the least important at the bottom. Columns correspond to individual models, and color gradients within each column indicate the most important (dark) to least important (light) variables.

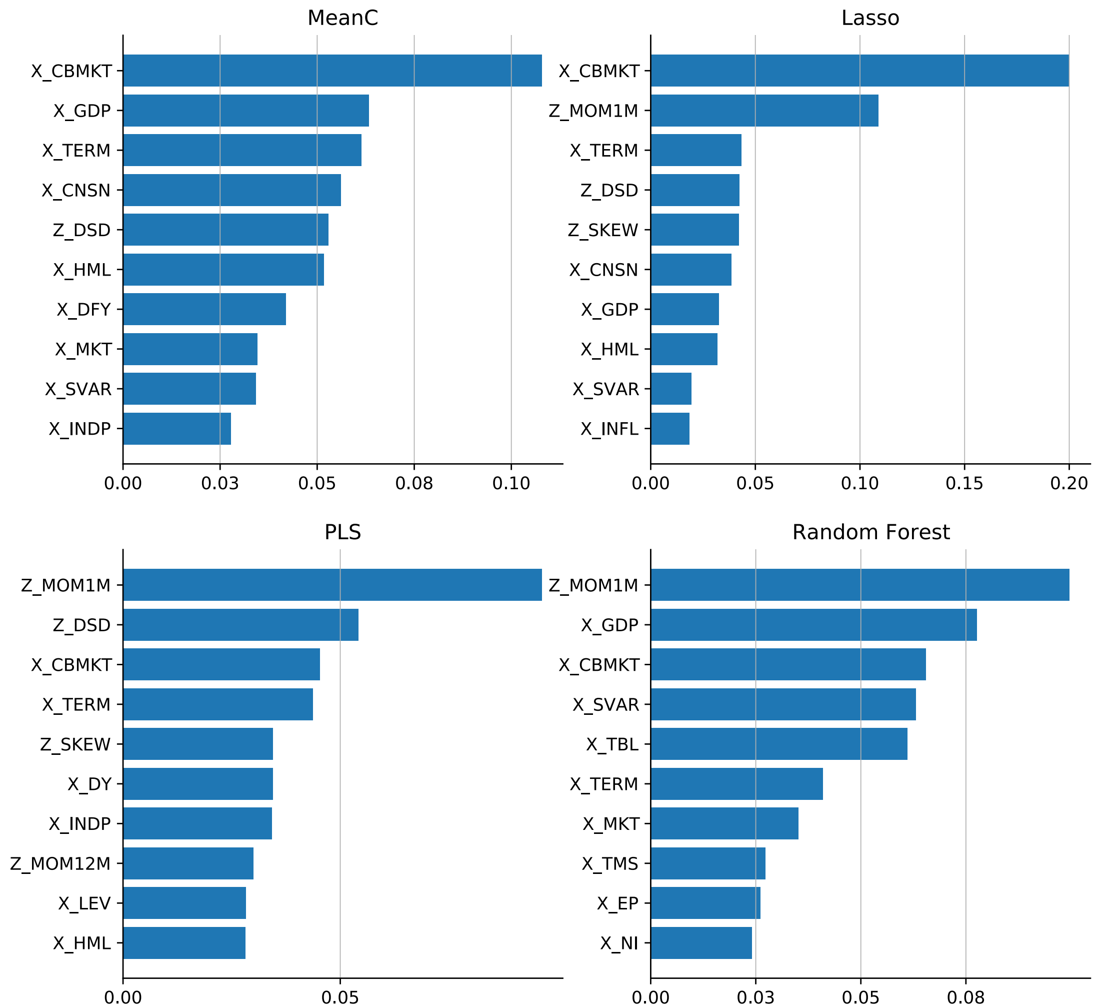

Figure 2: Relative Variable Importance by Model

This figure reports the relative variable importance for the 10 most influential variables in each model. We add the prefix "X" for aggregate predictors and "Z" for bond characteristics to distinguish the two groups of predictors. Variable importance is an average over all testing samples and corporate bonds. Variable importance within each model is normalized to sum to 1.

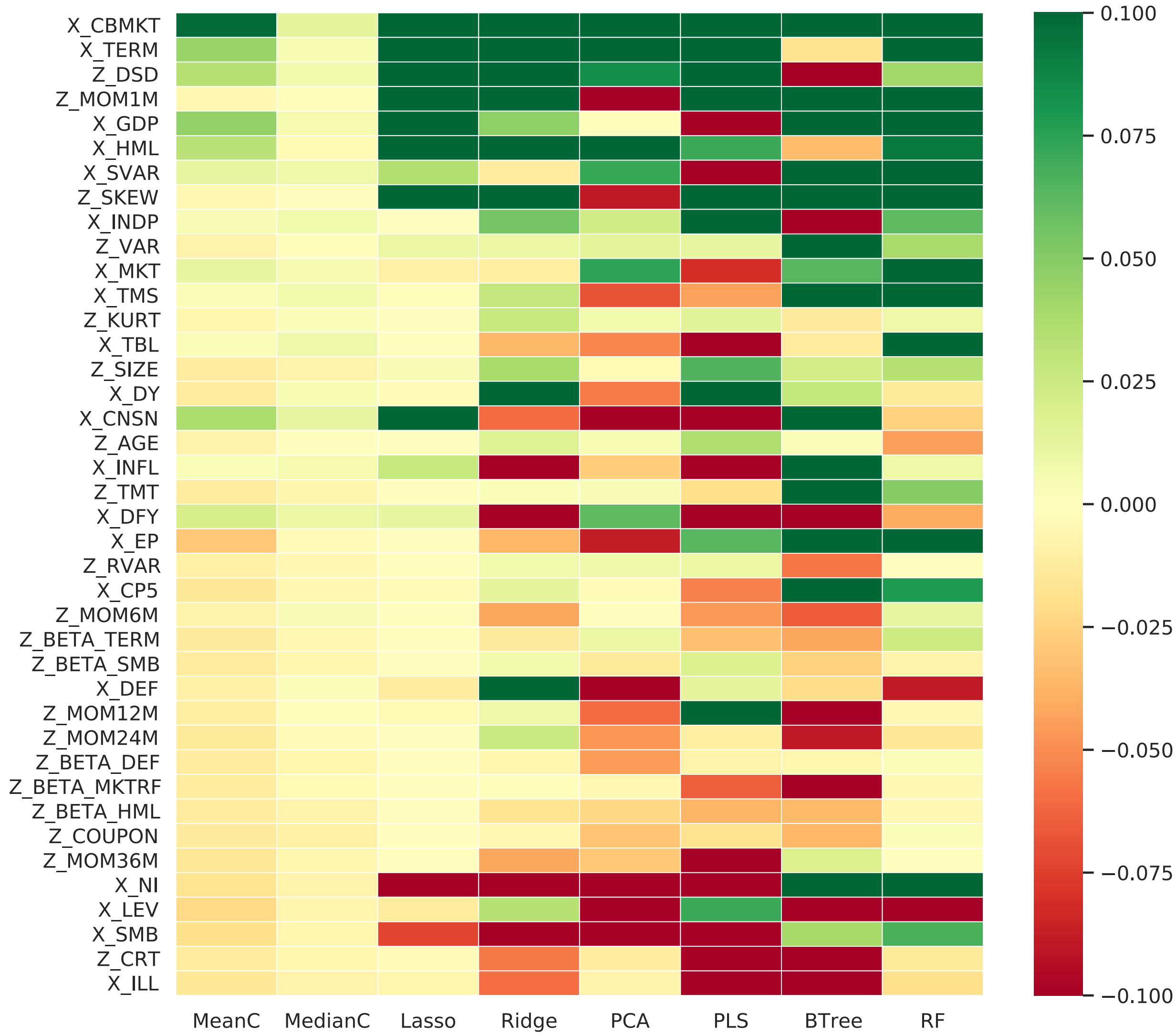

Figure 3: Heat Map of Variable Importance
Color should be used for this figure in print. This figure reports the variable importance of 40 predictors. We add the prefix "X" for aggregate predictors and "Z" for bond characteristics to distinguish the two groups of predictors. The order of predictors follows Figure 1. Moreover, the columns correspond to each model. A positive number (in green) means the variable improves OOS prediction, and a negative number (in red) means the predictor doesn't help OOS prediction.

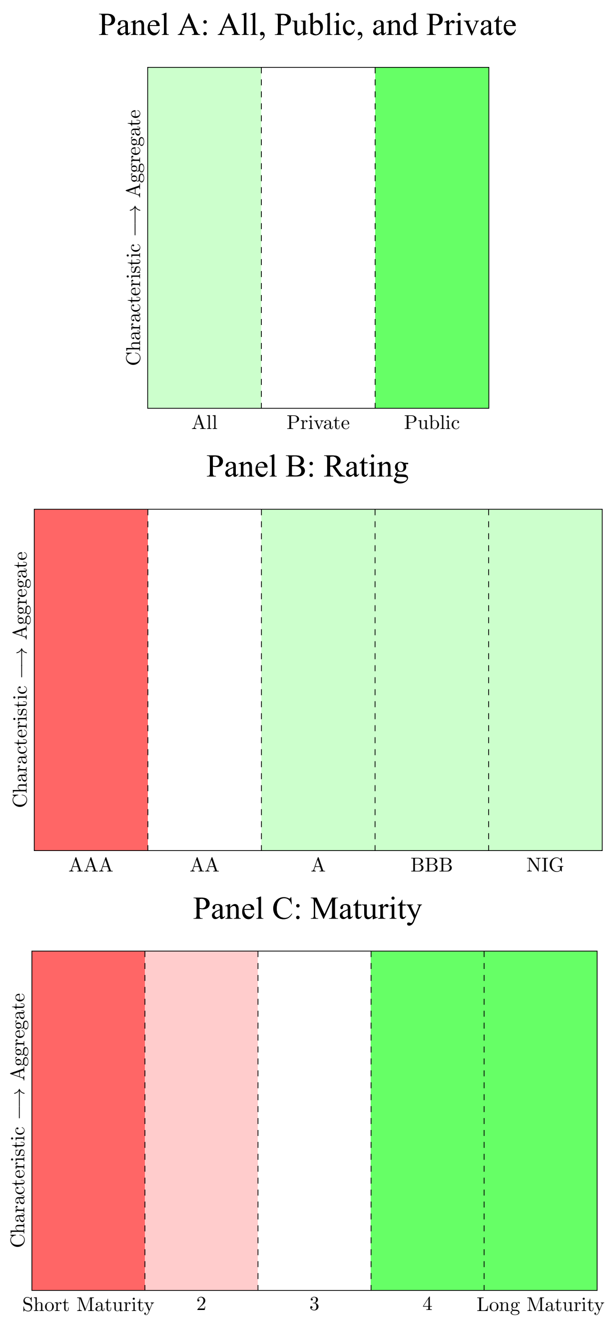

Figure 4: Relative Importance of Aggregate Predictor and Bond Characteristics
Color should be used for this figure in print. This figure shows the random forest model's relative difference between aggregate predictors and bond characteristics. **Green denotes an increase in importance, and red denotes a decrease from bond characteristics to aggregate predictors.** A deep color means the magnitude of the difference is large, and a light color means the difference is small. In Panel A, we plot all bonds and public and private groups. In Panel B, we plot different rating groups from AAA to non-investment-grade bonds. In Panel C, we plot different maturity quintile groups.

Figure 5: Relative Aggregate Predictor Importance

Color should be used for this figure in print. This figure displays the changes in aggregate predictor importance for the random forest model. **Green denotes an increase in importance, and red denotes a decrease**. The ordering of the aggregate predictors corresponds to their importance for the whole sample. We only report the 10 most important aggregate predictors for brevity, as listed in Table B.2. In Panel A, we plot the change from public to private bonds. In Panel B, we plot the change in aggregate predictor importance when moving from the AAA to non-investment-grade bonds. In Panel C, we plot the change from the short-maturity quintile to the long-maturity quintile.

Table 1: Summary Statistics

The sample includes 753,274 monthly return observations of 22,747 unique corporate bonds from January 1976 to September 2020. We report TRACE and NAIC together because they are both transaction-based data, and a large proportion of NAIC observations are covered by TRACE.

Panel A: Descriptive Statistics

	All	Lehman	DataStream	TRACE&NAIC	Public	Private
Bond-month observations	753,274	182,931	20,413	549,930	273,890	479,384
Start Year	1976	1976	1990	1993	1976	1976
End Year	2020	1998	2008	2020	2020	2020
% of Public Bond	36.36	43.30	5.32	34.99	100	0
% of Private Bond	63.64	56.70	94.68	65.01	0	100
% of IG	85.23	88.06	77.48	84.49	84.50	85.65
% of NIG	14.77	11.94	22.52	15.51	15.50	14.35
Return - mean (%)	0.51	0.78	0.51	0.41	0.51	0.49
Return - median (%)	0.39	0.69	0.42	0.28	0.43	0.37
Excess return - mean (%)	0.20	0.19	0.20	0.21	0.21	0.18
Excess return - median (%)	0.12	0.16	0.12	0.10	0.16	0.09
Rating - mean	5.60	5.58	7.46	5.53	6.02	5.35
Rating - median	5	5	7	5	5	5
Duration - mean (years)	5.61	5.37	8.91	5.58	5.83	5.49
Duration - median (years)	4.82	5.01	9.48	4.57	5.09	4.65
Age - mean (years)	9.00	17.21	6.78	6.10	9.11	8.93
Age - median (years)	5.63	18.94	6.34	4.39	5.80	5.54
Amt out. - mean ($ million)	1024	67	86	1155	515	1152
Amt out. - median ($ million)	130	25	10	150	200	100

Panel B: Sample Distribution (%) By Rating & Maturity

	AAA	AA	A	BBB	NIG	Public	Private	All
Maturity								
1	2.00	2.36	5.13	2.66	1.05	4.19	9.02	13.20
2	1.58	2.26	4.80	2.49	0.98	4.04	8.06	12.10
3	1.14	1.83	4.02	2.19	0.91	3.57	6.52	10.09
4	1.14	1.74	3.98	2.16	0.88	3.60	6.30	9.90
5	0.67	1.08	2.61	1.65	0.79	2.54	4.25	6.79
6	0.65	0.97	2.50	1.61	0.78	2.53	3.99	6.52
7	0.49	0.87	2.17	1.41	0.63	2.22	3.36	5.58
8	0.48	0.85	2.15	1.45	0.56	2.22	3.27	5.49
9	0.46	0.83	2.19	1.61	0.57	2.39	3.27	5.66
10	0.13	0.43	0.90	0.81	0.29	0.98	1.57	2.56
>10	1.69	2.64	7.47	7.56	2.74	8.09	14.02	22.11
All	10.43	15.87	37.92	25.60	10.18	36.36	63.64	100.00

Table 2: Return Predictive Evidence

This table reports the main results for corporate bond return predictability from 1996 to 2020. We report for eight predictive models, three factor models, and the naive zero prediction. The predictive models are introduced in Section 2.1 and Appendix A. The factor models are BS3 (equity market factor, corporate bond market factor, and treasury bond market factor) Bekaert and De Santis (2021), FF3 (MKT, DEF, and TERM) Fama and French (1993), and FF5 (MKT, SMB, HML, DEF, and TERM) Fama and French (1993). The left panel presents $\bar{R}^2_{OOS}$, as introduced in Equation (5), and the corresponding Fama-MacBeth *t*-statistics. The right panel presents R^2_{OOS} (%) as introduced in Equation (3). 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
MeanC	0.51***	0.02	0.39***	0.49	0.15	0.35
MedianC	0.04	-0.36	-0.03	-0.03	-0.13	-0.07
Lasso	3.50***	2.27***	3.06***	4.01	2.34	3.35
Ridge	3.06***	1.96***	2.72***	3.69	1.87	2.97
PCA	0.60	0.76**	0.67*	0.98	1.02	0.99
PLS	1.69***	1.13*	1.69***	2.64	1.99	2.38
BTree	4.78***	1.61**	3.60***	4.88	-0.39	2.80
RF	4.71***	2.28***	3.94***	4.34	2.22	3.50
BS3	0.00	-0.87*	-0.15*	0.07	-0.11	0.03
FF3	-0.82*	-0.16	-0.75*	-0.24	0.31	-0.13
FF5	-1.75***	-0.88**	-1.68***	-0.85	-0.02	-0.69
Zero	-0.92***	-1.32***	-0.95***	-0.27	-0.24	-0.26

Table 3: Return Predictive Evidence for Rating Groups

This table reports the return predictive evidence of five rating groups for the sub-periods of 1996-2007, 2008-2009, and 2010-2020. The table format follows Table 2.

	$\bar{R}^2_{OOS}$					R^2_{OOS}				
	AAA	AA	A	BBB	Junk	AAA	AA	A	BBB	Junk
Panel A: 1996-2020										
MeanC	0.15	0.35*	0.61***	0.62***	0.02	0.19	0.14	0.52	0.65	0.15
MedianC	-0.17	-0.08	0.16	0.12	-0.36	-0.36	-0.45	-0.02	0.18	-0.13
Lasso	2.85***	2.28***	3.11***	3.51***	2.27***	4.24	3.71	4.16	3.86	2.34
Ridge	1.88***	1.43*	2.64***	3.28***	1.96***	3.92	2.28	3.78	3.89	1.87
PCA	-0.55	-0.63	0.31	1.21**	0.76	0.05	-0.62	1.18	1.51	1.02
PLS	-0.11	-2.27*	0.83	2.10**	1.13*	1.77	0.07	2.84	3.44	1.99
BTree	5.47***	2.92***	4.54***	4.34***	1.61**	6.41	3.48	4.99	4.60	-0.39
RF	4.34***	4.37***	4.76***	4.52***	2.28***	4.71	3.83	4.42	4.27	2.22
BS3	0.14	0.18	0.26	-0.17	-0.87*	0.09	-0.05	0.08	0.09	-0.11
FF3	-0.65	-1.18**	-0.72	-0.71	-0.16	-0.08	-0.46	-0.32	-0.14	0.31
FF5	-1.26**	-2.26***	-1.78***	-1.62***	-0.88**	-0.36	-1.14	-1.04	-0.69	-0.02
Zero	-0.34**	-1.19***	-0.77***	-1.02***	-1.32***	-0.27	-0.43	-0.22	-0.29	-0.24
Panel B: 1996-2007										
MeanC	0.02	-0.21	0.04	0.40***	0.71**	-0.10	-0.38	0.04	0.46	-0.06
Lasso	4.53***	3.73***	3.46***	3.20***	1.99***	5.43	4.51	4.04	2.66	2.09
PLS	4.53***	0.92	1.58	1.59	1.33**	5.65	1.50	1.85	1.11	1.60
RF	5.95***	5.19***	4.95***	4.24***	1.76***	6.28	5.00	5.16	3.71	1.75
Panel C: 2008-2009										
MeanC	0.35**	0.54	0.76	0.63	0.38**	0.86	0.65	0.81	0.79	0.41
Lasso	-2.37*	3.06	4.54**	5.13***	2.15**	-1.83	2.62	4.55	6.05	2.49
PLS	-13.09***	0.17	5.26	6.32**	1.32	-16.59	-1.58	5.11	8.06	2.33
RF	1.27	3.13	5.40***	5.43***	2.62	-1.52	0.62	2.29	4.28	2.30
Panel D: 2010-2020										
MeanC	0.25	0.87***	1.20***	0.85***	-0.85**	0.42	1.08	1.49	0.99	0.20
Lasso	-0.20	-1.12	1.47	3.12**	2.10*	5.09	2.89	3.82	4.13	3.02
PLS	-5.79**	-7.30***	-1.69	1.17	0.26	3.75	-1.97	1.64	3.49	2.57
RF	1.87	3.06**	4.13***	4.56***	2.56	4.85	6.06	6.17	5.85	4.20

Table 4: Return Predictive Evidence for Maturity Groups

This table reports the return predictive evidence of five maturity groups for the sub-periods of 1996-2007, 2008-2009, and 2010-2020. The table format follows Table 2.

	$\bar{R}^2_{OOS}$					R^2_{OOS}				
	TMT1	TMT2	TMT3	TMT4	TMT5	TMT1	TMT2	TMT3	TMT4	TMT5
Panel A: 1996-2020										
MeanC	0.70***	0.56***	0.40***	0.33***	0.33***	0.35	0.36	0.38	0.38	0.32
MedianC	0.06	-0.02	-0.14	-0.10	0.02	-0.13	-0.17	-0.15	-0.03	-0.01
Lasso	0.95	2.76***	3.27***	3.25***	3.09***	3.22	3.83	4.15	3.29	2.86
Ridge	0.57	2.47***	3.10***	2.99***	2.61***	2.40	3.08	3.84	2.95	2.69
PCA	0.28	0.84	0.92*	0.65	0.46	0.23	0.99	1.32	1.09	1.02
PLS	-6.43	-0.52	2.00**	2.53***	2.55***	0.00	1.62	3.46	2.63	2.78
BTree	2.02**	3.64***	4.20***	3.07***	3.77***	3.16	3.90	4.23	1.82	2.19
RF	4.49***	4.62***	4.41***	3.84***	3.41***	3.74	4.27	4.25	3.18	2.97
BS3	-0.39	-0.77*	-0.58*	-0.37	0.14	0.04	-0.19	-0.12	0.07	0.18
FF3	-1.25	-1.06*	-0.84	-0.94**	-0.62	0.02	-0.14	-0.16	-0.14	-0.16
FF5	-2.37**	-2.07***	-1.66**	-1.85***	-1.57**	-0.54	-0.69	-0.70	-0.71	-0.71
Zero	-0.95***	-1.10***	-1.10***	-1.05***	-0.85***	-0.17	-0.27	-0.33	-0.26	-0.25
Panel B: 1996-2007										
MeanC	0.55**	0.35	0.25	0.24	0.19	0.07	0.03	0.05	0.09	0.06
Lasso	3.76***	3.56***	3.68***	2.84***	2.32***	4.02	4.30	4.20	2.62	1.97
PLS	0.39	1.29	2.62***	2.10**	1.90***	1.55	2.47	3.09	1.51	1.34
RF	4.59***	4.39***	4.71***	3.84***	3.11***	3.96	4.31	4.48	3.18	2.93
Panel C: 2008-2009										
MeanC	0.83**	0.76*	0.55	0.38	0.33	0.70	0.79	0.77	0.54	0.50
Lasso	4.14***	3.71**	3.84***	3.07**	2.07*	3.50	3.90	4.57	3.30	3.09
PLS	0.48	0.03	3.36	1.82	1.61	1.11	1.70	4.89	2.87	3.74
RF	5.00***	5.13**	4.73***	3.38*	2.43	2.97	3.54	3.57	1.93	1.81
Panel D: 2010-2020										
MeanC	0.78***	0.74***	0.50**	0.44**	0.48***	0.95	0.92	0.81	0.83	0.75
Lasso	-8.18***	-0.35	1.67	3.55***	4.23***	-2.58	1.52	3.32	5.08	5.12
PLS	-21.9***	-5.15	-0.24	3.01**	3.46***	-13.16	-2.45	2.34	5.24	5.24
RF	3.28**	4.31***	3.62***	3.81***	3.91***	5.02	5.65	4.57	5.60	5.52

Table 5: Predicting Private and Public Company Bonds

This table reports the time-series average $\bar{R}^2_{OOS}$ (%) with t-statistic significance, the difference in predictability between private and public bonds with t-statistic significance, and the R^2_{OOS} (%). For the public-private company bond predictability difference, we report the $\bar{R}^2_{OOS;\Delta}$ defined in Equation (6) with the two-sample Fama-MacBeth t-statistics. We find none of the differences are statistically significant, except in one case in Panels F. For t-statistics, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	$\bar{R}^2_{OOS}$		$\bar{R}^2_{OOS;\Delta}$	R^2_{OOS}	
	Private	Public	Difference	Private	Public
Panel A: All 1996-2020					
MeanC	0.38***	0.40***	-0.02	0.28	0.46
Lasso	3.01***	3.01***	0.01	3.46	3.18
PLS	1.57**	1.48*	0.09	2.43	2.32
RF	4.10***	3.59***	0.51	3.79	3.08
Panel B: All 1996-2007					
MeanC	0.28	0.24	0.04	0.02	0.14
Lasso	3.20***	3.16***	0.04	3.20	2.98
PLS	2.03***	2.02**	0.01	1.89	1.90
RF	4.27***	3.21***	1.05	4.04	2.81
Panel C: All 2008-2009					
MeanC	0.40	0.52	-0.12	0.55	0.67
Lasso	3.09***	2.90*	0.19	3.67	3.32
PLS	1.93	1.73	0.20	3.20	3.08
RF	3.87***	3.73*	0.14	2.45	2.42
Panel D: All 2010-2020					
MeanC	0.49***	0.52***	-0.02	0.76	0.90
Lasso	2.34**	2.17*	0.17	4.06	3.46
PLS	0.31	-0.36	0.67	3.02	1.79
RF	3.84***	3.65***	0.19	5.33	5.37
Panel E: IG 1996-2020					
MeanC	0.50***	0.56***	-0.05	0.38	0.69
Lasso	3.51***	3.42***	0.09	4.12	3.82
PLS	1.68**	1.31*	0.37	2.70	2.54
RF	4.68***	4.71***	-0.04	4.45	4.14
Panel F: NIG 1996-2020					
MeanC	0.01	-0.22	0.24	0.10	0.21
Lasso	1.99***	2.54***	-0.55	2.24	2.46
PLS	0.54	1.51*	-0.97*	1.92	2.06
RF	2.68***	1.91***	0.77	2.55	1.86
Panel H: TMT1 1996-2020					
MeanC	0.35***	0.36***	-0.01	0.28	0.46
Lasso	3.19***	3.11***	0.08	3.42	3.29
PLS	2.17***	1.95**	0.22	2.64	2.77
RF	4.00***	3.54***	0.46	3.69	3.16
Panel I: TMT5 1996-2020					
MeanC	0.35***	0.32***	0.03	0.28	0.39
Lasso	3.05***	3.22***	-0.17	2.70	3.08
PLS	2.58***	2.79***	-0.22	2.61	3.02
RF	3.57***	3.05***	0.52	2.97	2.98

Table 6: Investment Performance: Long-Short Strategy Returns

This table evaluates the investment gain of the forecast-implied long-short value-weighted portfolio returns. The performance measures 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 a test period from January 1996 to December 2020. In Panel B, the factor model follows Bai, Bali, and Wen (2019) (CBMKT, DRF, CRF, and LRF), with a test period from July 2004 to December 2020. For *t*-statistics, *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	All			IG			NIG			Private			Public		
	Mean	α	SR	Mean	α	SR	Mean	α	SR	Mean	α	SR	Mean	α	SR
Panel A: 1996-2020 Alpha on FF5															
MeanC	0.73	0.64***	1.21	0.66	0.62***	1.29	0.80	0.79***	0.63	0.80	0.72***	1.25	0.38	0.26*	0.58
Lasso	1.12	1.05***	2.37	1.10	1.05***	2.52	1.51	1.60***	1.17	1.20	1.14***	2.46	0.80	0.70***	1.40
PLS	1.20	1.16***	2.19	1.18	1.16***	2.44	1.60	1.61***	1.35	1.29	1.25***	2.22	0.83	0.76***	1.43
RF	1.86	1.83***	3.41	1.80	1.80***	3.73	1.70	1.55***	1.40	2.02	2.02***	3.64	1.19	1.09***	1.83
Panel B: 2004-2020 Alpha on BBW4															
MeanC	1.02	0.50***	1.58	0.98	0.57***	1.84	1.06	0.85**	0.77	1.13	0.60***	1.66	0.61	0.08	0.90
Lasso	1.23	0.93***	2.44	1.21	0.97***	2.65	1.41	1.20***	1.05	1.31	1.05***	2.57	0.87	0.47***	1.53
PLS	1.11	0.75***	1.97	1.12	0.86***	2.27	1.40	1.05***	1.18	1.23	0.90***	2.12	0.69	0.29**	1.20
RF	1.63	1.30***	3.09	1.55	1.35***	3.41	1.41	0.70**	1.18	1.78	1.53***	3.44	0.99	0.44***	1.48

Table 7: Investment Performance: Market-Timing Strategy Returns

This table reports the investment gain of the forecast-implied market-timing value-weighted strategy. The table format follows Table 6.

	All			IG			NIG			Private			Public		
	Mean	α	SR	Mean	α	SR	Mean	α	SR	Mean	α	SR	Mean	α	SR
Panel A: 1996-2020 Alpha on FF5															
MeanC	0.75	0.73**	0.42	0.40	0.34*	0.48	0.85	0.69*	0.45	0.87	0.82**	0.59	0.37	0.39	0.29
Lasso	1.33	1.52***	0.73	1.48	1.46***	0.97	1.47	1.67***	0.67	1.58	1.71***	0.94	0.50	0.65*	0.26
PLS	1.52	1.49***	0.96	1.72	1.69***	1.08	1.28	1.44***	0.67	1.32	1.22***	0.74	0.92	0.95***	0.65
RF	1.51	1.48***	1.72	1.49	1.49***	2.17	1.12	0.99***	0.84	1.59	1.56***	1.76	0.97	0.95***	1.28
Panel B: 2004-2020 Alpha on BBW4															
MeanC	0.76	0.51	0.56	0.32	0.15**	0.98	1.09	0.22	0.65	0.75	0.55	0.55	0.39	0.03	0.88
Lasso	1.17	1.59***	0.70	1.60	0.73**	1.17	1.46	1.50***	0.69	1.55	1.02***	1.13	0.10	0.84*	0.05
PLS	0.93	1.09***	1.16	1.23	0.96***	1.37	0.91	0.77**	0.71	0.71	0.98***	0.58	0.45	0.64***	0.71
RF	1.17	1.17***	1.43	1.07	1.11***	2.05	1.43	0.88***	1.17	1.28	1.34***	1.46	0.70	0.57***	1.17

Table 8: Investment Performance with a Subset of Predictors

This table reports the investment gain of machine learning investment strategies, with both aggregate predictors and bond characteristics, with only aggregate predictors and only bond characteristics. The only aggregate strategy in Panel A is not applicable, so we leave the numbers blank. The α is based on Fama-French five factors (MKT, SMB, HML, TERM, DEF) in Fama and French (1993) with a test period from January 1996 to December 2020. The table format follows Table 6.

	Aggregate+Characteristic			Only Aggregate			Only Characteristic		
	Mean	α	SR	Mean	α	SR	Mean	α	SR
Panel A: Long-Short Portfolio Returns									
MeanC	0.73	0.64***	1.21				0.73	0.64***	1.21
Lasso	1.12	1.05***	2.37				0.63	0.55***	1.56
PLS	1.20	1.16***	2.19				0.83	0.79***	1.63
RF	1.86	1.83***	3.41				1.37	1.30***	2.23
Panel B: Market-Timing Portfolio Returns									
MeanC	0.75	0.73**	0.42	0.29	0.27***	0.90	0.46	0.47	0.22
Lasso	1.33	1.52***	0.73	0.47	0.47***	1.55	0.35	0.33***	1.19
PLS	1.52	1.49***	0.96	0.40	0.39***	1.28	0.43	0.41***	1.30
RF	1.51	1.48***	1.72	0.39	0.38***	1.23	0.64	0.59***	1.75

Table 9: Portfolio Performance with Transaction Cost

This table evaluates the impact of transaction costs on machine-learning investment strategies. We report the turnover ratio, monthly returns, and Sharpe ratio for different transaction cost levels.

	Turnover (%)	Monthly Average Returns (%)					Annualized Sharpe Ratio				
		0 bp	20 bp	40 bp	60 bp	80 bp	0 bp	20 bp	40 bp	60 bp	80 bp
Panel A: Long-Short											
MeanC	64.21	0.73	0.60	0.47	0.34	0.21	1.21	1.00	0.79	0.57	0.36
Lasso	100.28	1.12	0.92	0.72	0.52	0.32	2.37	1.96	1.54	1.12	0.69
PLS	94.23	1.20	1.01	0.83	0.64	0.45	2.19	1.86	1.52	1.18	0.83
RF	141.11	1.86	1.57	1.29	1.01	0.73	3.41	2.89	2.37	1.85	1.33
Panel B: Market-Timing											
MeanC	35.32	0.75	0.68	0.61	0.54	0.47	0.42	0.38	0.34	0.30	0.26
Lasso	118.00	1.33	1.10	0.86	0.62	0.39	0.73	0.60	0.47	0.34	0.21
PLS	119.79	1.52	1.28	1.04	0.81	0.57	0.96	0.81	0.66	0.51	0.36
RF	105.42	1.51	1.30	1.09	0.88	0.67	1.72	1.49	1.25	1.01	0.77

Appendix A. Brief introduction to predictive models and design

A.1 Linear regression

The most used conditional expectation functional form in the empirical asset pricing literature is the predictive regression:

$$g(z_{it}, x_t; \theta) = [1, z'_{i,t}, x'_t]\theta. \quad \text{(A-1)}$$

If we fit a pooled regression for individual bonds, the number of observations is typically much higher than the number of predictors. Therefore, overfitting is not a serious concern, but the bias-variance trade-off remains an issue for OOS prediction.

A.2 Forecast combination

The combination forecast of $R_{i,t+1}$ is usually the weighted average of J individual forecasts:

$$\hat{R}^c_{i,t+1} = \sum_{j=1}^{J} \omega_{i,j,t} \hat{R}_{i,j,t+1}, \quad \text{(A-2)}$$

where $\hat{R}^c_{i,t+1}$ 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 predictor j, and $\omega_{i,j,t}$ is the weight to combine individual forecasts. 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 regard these robust combination forecasts as the benchmark in forecast performance evaluation.

A.3 Dimension reduction

We consider two classic dimension reduction techniques: principal component analysis regression (PCA) and partial least squares (PLS), which are commonly used for latent factor models in empirical asset pricing. One can use a low-dimension version of linearly transformed predictors to construct a predictive model and solve the bias-variance trade-off problem.

PCA and PLS regression consist of a two-step procedure. The first step of PCA combines predictors with a small set of linear combinations that best preserve the covariance structure of these predictors. The first K principal components are used in multiple regressions in the second step. The first step of PLS combines predictors with a small set of linear combinations that best preserve the covariance between predictors and outcomes. The first K components are used in multiple regressions in the second step. In contrast to PCA, the PLS objective seeks K linear combinations of the independent variables with a maximal predictive correlation with the dependent variable. K is set to 3 in our applications for both PCA and PLS.

A.4 Regularized predictive regressions

Regularized linear predictive regressions, such as lasso and ridge, are also commonly used in machine learning finance. By adding a penalty over ordinary linear regression, these two linear machine learning methods preserve the interpretability of linear models. Compared with PCA and PLS, the advantage of these two models is that they preserve the predictors without transforming them and thus keep the original interpretation.

Lasso and ridge share similar loss functions but have different regularization effects. Lasso adds an L1 norm penalty on the sum of residual squares, whereas ridge adds an L2 norm penalty. On the one hand, the 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 the penalty weight, with a larger penalty weight imposing more shrinkage on the coefficients. We discuss the selection of tuning parameters in Section A.6.

A.5 Ensemble tree models

Regression trees are among the most popular machine learning algorithms. This approach is an alternative to nonlinear regressions to partition the space into smaller regions where the interaction is 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 tree and random forest.

The gradient-boosted 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 weak forecasts from simple trees. The predictions by individual trees could be weak, but the prediction from adding up all these tree forecasts becomes 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 and adopts the bootstrap aggregating scheme. One draws B number of data bootstrap samples and fits B different trees. One usually chooses a random subset of predictors for each tree and fits a tree model with the bootstrap sample. Again, the predictions by individual trees could be weak, but the average of all B numbers of forecasts is strong. Averaging the individual forecasts reduces the overfitting in individual bootstrap samples and provides a stable forecast. Tuning parameters and model selection for these two ensemble tree methods are discussed in Section A.6.

A.6 Deterministic cross-validation

Many machine learning models require tuning parameters. We adopt a deterministic fivefold cross-validation scheme as illustrated in Figure B.1. Standard randomized cross-validation is initially designed for independent observations. Our deterministic cross-validation design preserves the continuity of time-series predictability in consecutive periods.

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 four of the five folds and validate using the remaining one fold, resulting 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.

Concerning the predictors' heterogeneity, we standardize each predictor before cross-validation. In a given year K, we calculate the sample mean $\hat{\mu}_K$ and standard deviation $\hat{\sigma}_K$ using each predictor's data in the past 20 years. We then standardize the predictors in the sample of the past 20 years and current year K with $\hat{\mu}_K$ and $\hat{\sigma}_K$. The test sample predictors are standardized by the same sample mean and standard deviation to allay forward-looking concerns.

Appendix B. Additional tables and figures

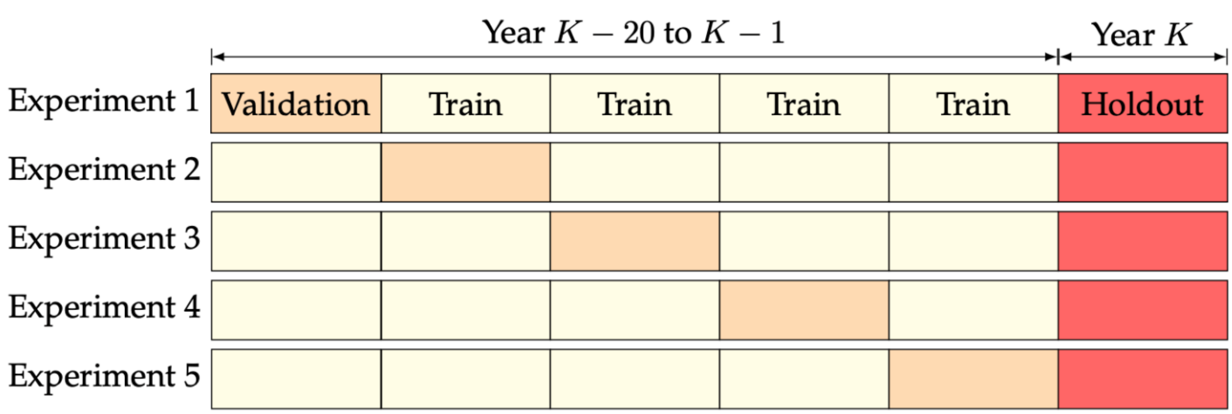

Figure B.1: Deterministic Five-Fold Cross-Validation
This figure demonstrates the deterministic five-fold cross-validation scheme. At the end of each year, we re-estimate the models using data for the past 20 years. Specifically, the deterministic design divides the sample into five consecutive parts. The rest of the cross-validation procedures follow the standard approach in Gu, Kelly, and Xiu (2020).

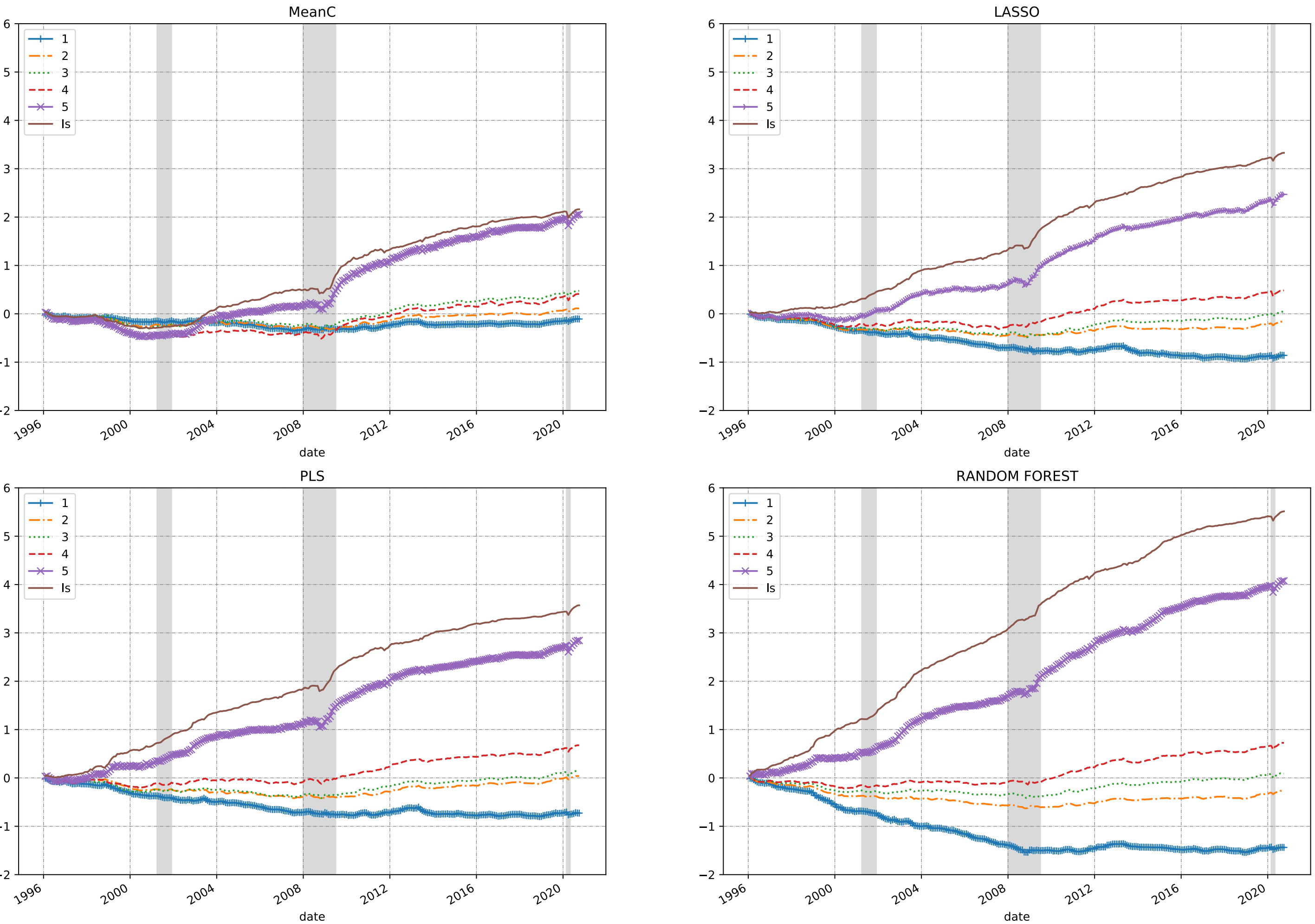

Figure B.2: Investment Performance of Long-Short Strategy: Value-Weighted Portfolio Return
This figure shows the cumulative returns plot of the value-weighted long-short strategy and sorted portfolios. Shadowed areas are recession periods of the NBER US Business Cycle Indicator.

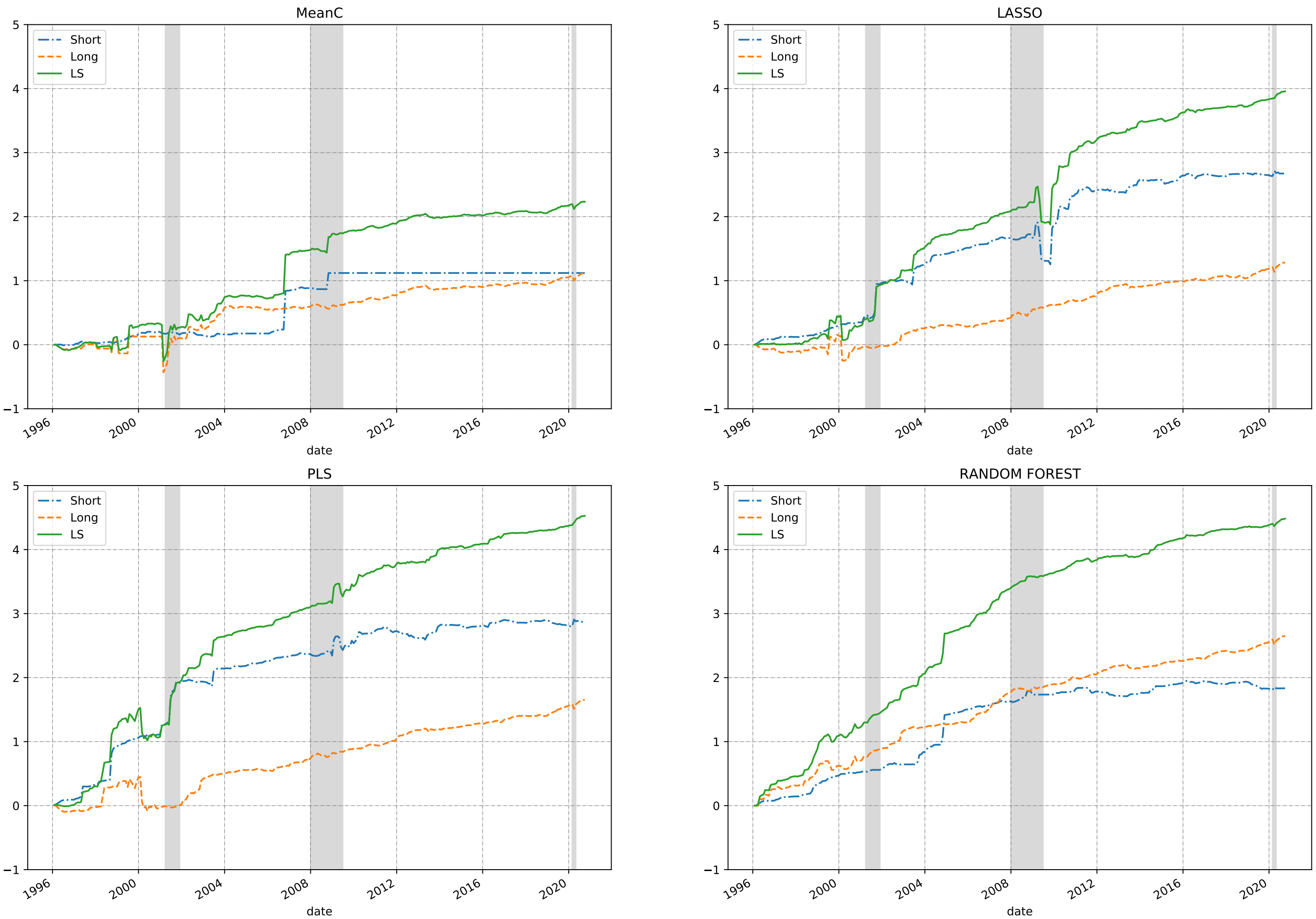

Figure B.3: Investment Performance of Market-Timing Strategy: Value-Weighted Portfolio Return
This figure shows the cumulative-returns plot of the value-weighted market-timing strategy. The table format follows Figure B.2.

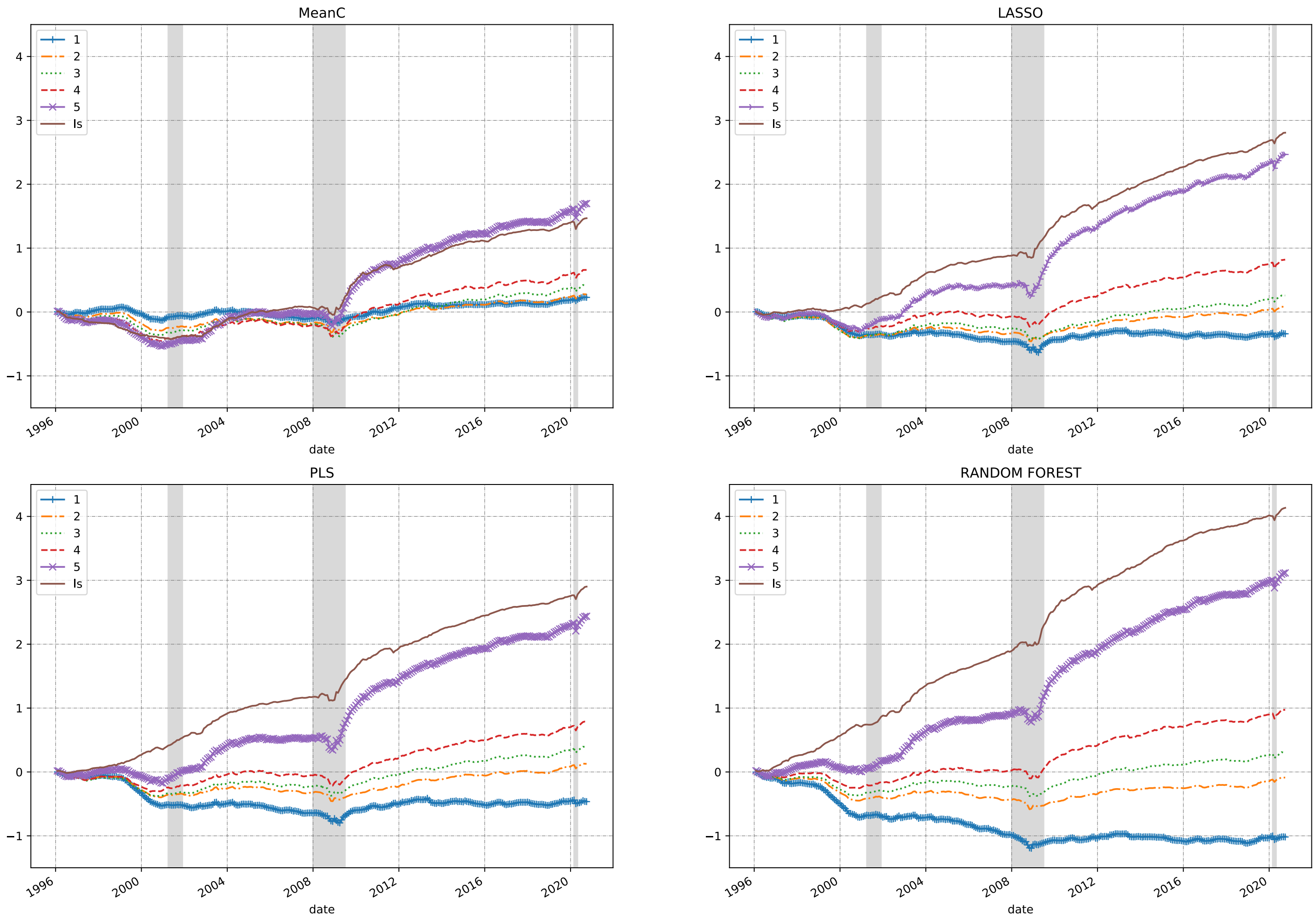

Figure B.4: Investment Performance of Long-Short Strategy: Equal-Weighted Portfolio Return
This figure reports the cumulative return of the equal-weighted long-short strategy. The value-weighted version is in Figure B.2.

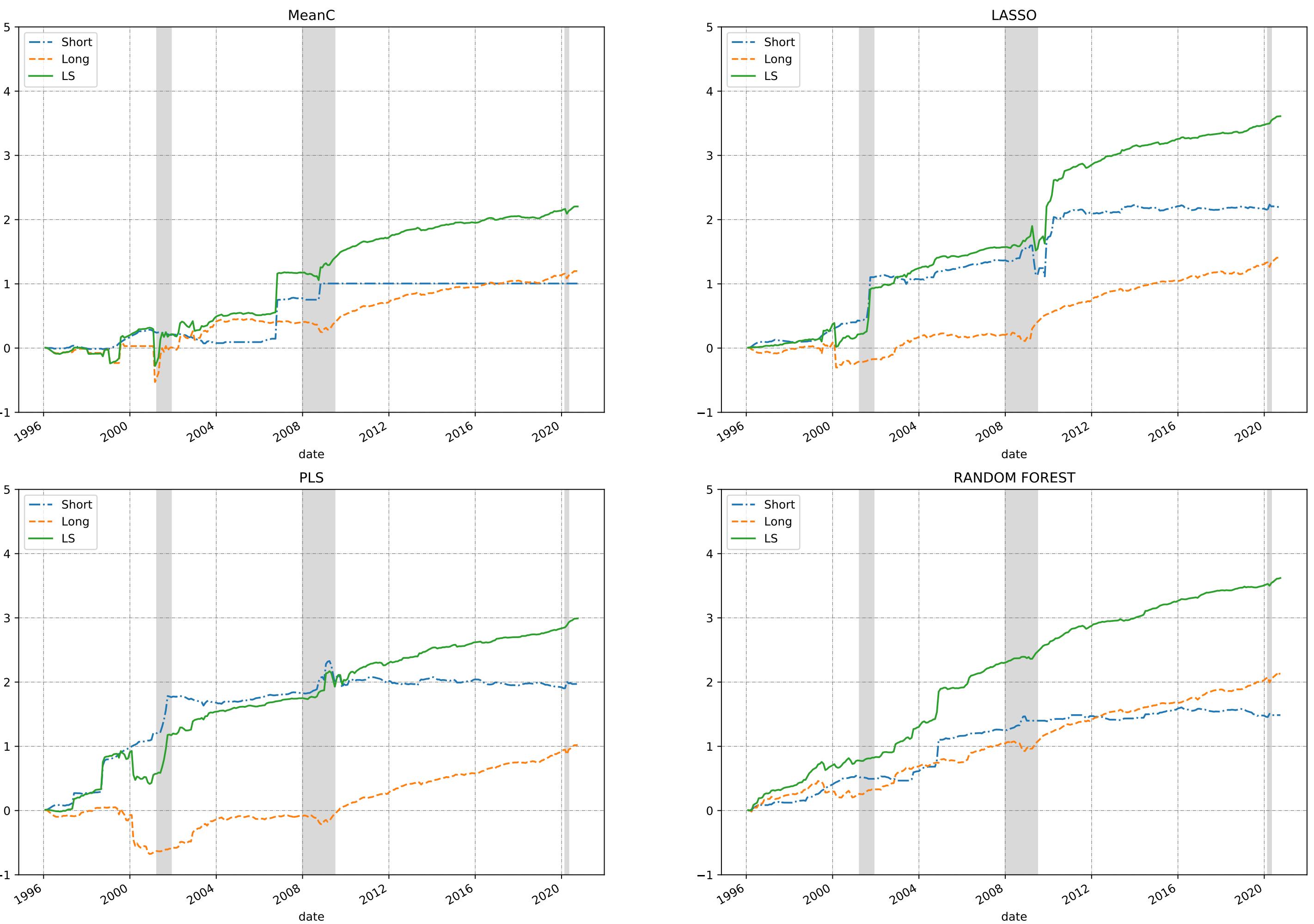

Figure B.5: Investment Performance of Market-Timing Strategy: Equal-Weighted Portfolio Return
This figure reports the cumulative return of the equal-weighted market-timing strategy. The value-weighted version is in Figure B.2.

Table B.1: Predictor List

Acronym	Description	Details
		Aggregate Predictors
Bond market variable		
TBL	3-month treasury bill rate	Download from Fed. St. Louis, Welch and Goyal (2008)
GDP	GDP growth rate	Annual percentage growth rate in GDP, Giesecke et al. (2011)
INDP	Industry production growth rate	Annual percentage growth rate in industrial production, Giesecke et al. (2011)
CNSN	Consumption growth rate	Annual percentage growth rate in per capita real personal consumption, Giesecke et al. (2011)
CBMKT	Corporate bond market return	Value-weighted corporate bond market return, Bai, Bali, and Wen (2019)
TERM	TERM factor	Long-term government bond return minus the one-month Treasury bill rate, Fama and French (1993)
DEF	Default factor	Long-term corporate bond return minus long-term government bond return, Fama and French (1993)
INFL	CPI index	Annual percentage growth rate in CPI index, Welch and Goyal (2008)
CP5	Forward factor	Codes of Monika Piazzesi, 5-year specification, Cochrane and Piazzesi (2005)
TMS	Term spread	Long-term yield on government bonds (from Ibbotson Associates), Welch and Goyal (2008)
DFY	Default yield spread	Yield of BAA- corporate bond minus yield of AAA corporate bonds, Welch and Goyal (2008)
Equity market variable		
DP	Dividend-to-price	S&P500 index dividend-to-price, Welch and Goyal (2008)
EP	Earnings-to-price	S&P500 index earnings-to-price, Welch and Goyal (2008)
NI	Net equity issuance	S&P500 index net equity issuance, Welch and Goyal (2008)
LEV	Leverage	S&P500 index leverage, Welch and Goyal (2008)
SVAR	Stock variance	S&P500 index variance, Welch and Goyal (2008)
MKTRF	Market factor	Download from Kenneth French website, Fama and French (1993)
SMB	Size factor	Download from Kenneth French website, Fama and French (1993)
HML	Value factor	Download from Kenneth French website, Fama and French (1993)
ILL	Pastor-Stambaugh illiquidity	Download from Stambaugh website, Pastor and Stambaugh (2003)
		Corporate Bond Characteristics
Fundamental		
CRT	Credit rating	From FISD
DUR	Duration	From FISD
TMT	Time-to-maturity	From FISD
AGE	Time-from-issuance	From FISD
SIZE	Amount outstanding	From FISD
Return-distribution		
STR	Short-term reversal	Lag 1-month return, Bai, Bali, and Wen (2019)
MOM6M	6-month momentum	Lag 2-month to lag 6-month cumulative return, Jostova et al. (2013)
MOM12M	12-month momentum	Lag 2-month to lag 12-month cumulative return
LTR2Y	2-year long-term reversal	Lag 13-month to lag 24-month cumulative return
LTR3Y	3-year long-term reversal	Lag 13-month to lag 36-month cumulative return, Bali, Subrahmanyam, and Wen (2021)
VAR	Variance	Variance of returns of the past 36 months, Bai, Bali, and Wen (2016)
DSD	Downside risk	5% VaR of returns of the past 36 months, Bai, Bali, and Wen (2019)
SKEW	Skewness	Skewness of returns of the past 36 months, Bai, Bali, and Wen (2016)
KURT	Kurtosis	Kurtosis of returns of the past 36 months, Bai, Bali, and Wen (2016)
Covariance on risk factors		
BETA_MKT	Multiple regression beta of a five-factor model, Fama and French (1993)	
BETA_SMB	Multiple regression beta of a five-factor model, Fama and French (1993)	
BETA_HML	Multiple regression beta of a five-factor model, Fama and French (1993)	
BETA_DEF	Multiple regression beta of a five-factor model, Fama and French (1993)	
BETA_TERM	Multiple regression beta of a five-factor model, Fama and French (1993)	
RVAR	Residual variance in the multiple regression of a five-factor model, Fama and French (1993)	

Table B.2: Relative Variable Importance of Aggregate Predictors

The table presents the relative variable importance of aggregate predictors for each predictive model. The sum of variable importance is normalized to 1. All values are in percentage.

	MeanC	MedianC	Lasso	Ridge	PCA	PLS	BTree	RF
CBMKT	16.78	9.97	36.33	12.62	13.88	10.64	13.67	11.16
TERM	9.55	5.60	7.87	8.64	8.57	10.28	3.86	6.99
SVAR	5.31	8.12	3.54	4.48	5.92	2.77	13.38	10.76
GDP	9.85	6.72	5.93	5.06	4.77	3.04	5.23	13.22
HML	8.04	2.09	5.81	7.43	10.46	6.63	3.75	3.95
CNSN	8.71	9.35	7.00	3.99	3.18	3.67	4.94	2.30
TBL	4.23	7.53	2.72	4.24	3.96	3.87	3.89	10.41
MKT	5.38	6.30	2.49	4.49	5.97	4.43	4.40	6.00
INDP	4.32	7.03	2.72	5.13	5.15	8.06	3.23	3.51
TMS	4.20	7.05	2.70	4.86	3.71	4.97	4.74	4.66
DY	2.32	5.75	2.67	5.71	3.91	8.11	4.17	2.46
DFY	6.52	8.54	2.99	3.30	5.76	2.97	0.00	2.09
INFL	4.22	6.43	3.36	3.28	4.35	2.48	5.08	2.78
EP	0.00	2.60	2.72	4.23	3.39	6.53	5.59	4.45
DEF	2.72	4.51	2.46	6.05	2.47	5.79	3.84	1.41
CP5	1.98	1.06	2.68	4.72	4.75	4.82	4.65	3.76
ILL	1.99	0.00	2.59	4.00	4.66	4.03	2.25	2.39
LEV	0.91	0.37	2.44	4.92	0.00	6.63	3.30	0.00
SMB	1.33	0.91	0.98	2.84	2.86	0.28	4.24	3.59
NI	1.62	0.08	0.00	0.00	2.29	0.00	5.78	4.10