Harness Machine Learning with Carry *

Fang Qiao, Zhan Shi, and Yicheng Zhu †

May 31, 2025

Abstract

This paper proposes a hybrid machine learning (ML) framework for fixed-income securities, by integrating ML-based forecasts of price returns with bond carry to form composite expected return estimates. Using a comprehensive dataset of corporate bonds over the 1992–2019 period, we illustrate that the proposed ML estimators deliver superior forecasting accuracy and dominate conventional ML models applied directly to total returns. This dichotomy highlights bond markets' efficiency in pricing carry-related risks and ML's incremental value in capturing deviations from yield-implied expectations. High-yield bonds exhibit the greatest return predictability, driven by nonlinear credit risk and liquidity shocks, whereas carry strategies remain superior for bonds issued by unlisted firms. Portfolio tests reveal that long-short strategies based on pure ML approaches generate gross monthly alphas of 1.74% per month, but transaction costs erode 60–80% of gains due to high turnover. In contrast, carry-driven ML strategies sustain net returns of 1.15% with lower turnover. Our results highlight the necessity of integrating asset-class-specific structure into ML frameworks and caution against direct adoption of equity-style strategies in fixed income.

JEL Classification: G12, G17, C38, C45, C53

Keywords: Corporate bond returns, Machine learning, Return predictability, Carry

*We would like to thank Stefano Giglio, Amit Goyal, Jingzhi Huang, Kewei Hou, Jiacui Li, Baozhong Yang and Jianfeng Yu; seminar participants at CUHK-Shenzhen, Fudan University, Peking University, and Tsinghua University for helpful comments and suggestions.

†Qiao, China School of Banking and Finance at University of International Business and Economics, 10 Huixindongjie, Chaoyang District, Beijing 100029, China, qiaofang@uibe.edu.cn; Zhan, PBC School of Finance, Tsinghua University, Beijing, 100083, China; shizh@pbcsf.tsinghua.edu.cn. Zhu, PBC School of Finance, Tsinghua University, Beijing, 100083, China.

1 Introduction

The rapid adoption of machine learning (ML) in empirical asset pricing has equipped researchers with novel tools to address high dimensionality, nonlinearities, and interaction effects inherent in return prediction. Recent advances in this field show that ML methods significantly outperform traditional linear models in inferring risk premium on individual securities and, more importantly, improvements in predictive accuracy translate into economically meaningful portfolio gains. On the other hand, machine-learning-based investing is still in its infancy for fixed-income securities, where factor strategies still comprise a limited part of the overall investment strategies. While machine learning methods shown fruitful in other asset classes seem equally applicable to fixed-income markets, bond returns present unique structural characteristics. Specifically, the inherent predictability embedded in yield curve dynamics—formalized through the concept of carry—provides a natural starting point for bond return prediction. Indeed, Koijen, Moskowitz, Pedersen, and Vrugt (2018) show that carry strategies perform well in government bond and credit markets. In this paper, we aim to build augmented machine learning estimators by incorporating this inherent component of bond returns, and empirically test its effectiveness for an important class of fixed-income assets, namely corporate bonds.

Bond carry is the expected return on a bond when the yield curve stays the same. To motivate our analysis, consider the h-period excess return on a m-period discount bond,

$$
\begin{aligned}
rx_{t,t+h}^{(m)} &= -(m-h)y_{t+h}^{(m-h)} + my_t^{(m)} - hy_t^{(h)} \\
&= \underbrace{-(m-h)E_t\left(y_{t+h}^{(m-h)}\right) + my_t^{(m)} - hy_t^{(h)}}_{\text{expected total return}} + \underbrace{(m-h)\left[E_t\left(y_{t+h}^{(m-h)}\right) - y_{t+h}^{(m-h)}\right]}_{\text{unexpected price shock}} \\
&= \underbrace{-(m-h)y_t^{(m-h)} + my_t^{(m)} - hy_t^{(h)}}_{\text{carry}} + \underbrace{(m-h)\left[y_t^{(m-h)} - E_t\left(y_{t+h}^{(m-h)}\right)\right]}_{\text{expected price return}} \\
&\quad + \underbrace{(m-h)\left[E_t\left(y_{t+h}^{(m-h)}\right) - y_{t+h}^{(m-h)}\right]}_{\text{unexpected price shock}}
\end{aligned}
$$

where y_t denotes log yield at time t. Therefore, the predictable component of a bond's excess return equals its carry plus its expected price appreciation. That carry is a model-free characteristic directly measurable ex ante from the yield curve makes it special. Expected price return, by contrast, need to be inferred using an asset pricing model or in a mode-free way. Building on this insight, we focus on machine learning forecasts of price returns and combine them with bond carry to form a composite estimator of total returns. In particular, we are interested in the horse race between this carry-based estimator and an alternative one that directly apply machine learning methods to total return prediction.

Corporate bonds represent a cornerstone of global financial markets, with over $10 trillion outstanding in the U.S. alone, yet their return dynamics remain less understood than those of equities. While extensive literature examines cross-sectional predictability in stock returns, corporate bond markets present unique challenges: illiquidity, heterogeneous credit risk, and sparse transaction data, particularly for older issuances. We test the empirical performance of alternative machine learning approach to predicting corporate bond returns, by leveraging a comprehensive dataset spanning over 1.26 million U.S. bond-month observations from 1992 to 2019. In particular, we consider 60 bond-specific characteristics and 153 firm/equity characteristics to capture both security-level and issuer-level determinants of bond returns.

First, we demonstrate that the carry-driven ML strategy substantially outperform the conventional one for all ML algorithms, achieving out-of-sample R^2 values up to 3.68% for bonds issued by listed firms. Even when predicting total excess returns, a simple yield-implied carry benchmark dominates all ML approaches, with an R^2 of 2.95% versus 2.24% for the best ML model. This result underscores the efficiency of bond markets in pricing carry-driven compensation but highlights ML's edge in capturing deviations from yield-implied expectations.

Second, we document stark differences in predictability across bond types. For high-yield bonds, ML models generate superior price return forecasts ($R^2 = 5.13\%$) compared to investment-grade issues, reflecting the greater role of nonlinear credit risk dynamics and liquidity shocks in speculative-grade markets. Conversely, yield-based benchmarks remain dominant for private firms, where equity data unavailability limits ML's informational advantage.

Third, we assess economic value through portfolio strategies. While ML-based long-short decile portfolios yield gross monthly alphas of 1.74%, transaction costs erode 60–80% of these gains due to high turnover. In contrast, yield-implied strategies exhibit lower turnover and sustain net returns of 1.15%, suggesting that much of ML's statistical predictability reflects compensation for liquidity risk rather than exploitable mispricing.

Our findings have critical implications for both theory and practice. They affirm that corporate bond markets efficiently price yield-driven risk premia but leave room for ML to enhance price return forecasts, particularly in high-yield segments. For investors, the results caution against direct adoption of equity-style ML strategies in bonds unless coupled with transaction cost mitigation. Methodologically, we advance the literature by integrating carry decomposition into ML frameworks—a approach that could extend to other fixed-income assets.

This paper contributes to the broad research agenda that seeks to characterize and understand differences in expected returns across assets. A recently emerging literature employs a gamut of machine learning methods to extract from a "zoo" of characteristics the information most relevant to individual security returns, including stock returns (Gu, Kelly, and Xiu, 2020; Chen, Pelger, and Zhu, 2023; Leippold, Wang, and Zhou, 2021), bond returns (Bianchi, Bchner, and Tamoni, 2021; Bali, Goyal, Huang, Jiang, and Wen, 2021a; Feng, He, Wang, and Wu, 2025), option returns (Bali, Beckmeyer, Moerke, and Weigert, 2023; Goyenko and Zhang, 2021), and currency returns (Filippou, Rapach, Rapach, Taylor, and Zhou, 2021). Most all of these studies show that machine learning enables more accurate inferences of asset risk premiums and brings sizable economic gains to investors. Our results point to several potential pitfalls in applying these techniques to certain asset markets, which are dominated by institutional investors and suffers great market illiquidity. The impact of transaction costs on the economic significance of return predictability is discussed in Chordia, Goyal, Nozawa, Subrahmanyam, and Tong (2017). We provide complementary evidence that, as an increasing number of characteristics are included in the formation of return forecasts, trading costs of corresponding high-minus-low strategy increase dramatically. While machine learning offers valuable insights into the complex dynamics of bond prices, our findings suggest that its practical application is constrained by market frictions and the efficiency of yield-based risk premiums.

On the other hand, the bond carry serves as a natural benchmark in fixed income markets and related trading strategies are widely used by practitioners since 1990s (Ilmanen and Sayood, 2002). This paper contributes to this strand of literature by introducing ML frameworks that explicitly account for bond carry. In this sense, our results also complement and extend the finding of Koijen et al. (2018) that carry is a strong positive predictor of returns in many asset classes. Our findings affirm that corporate bond markets efficiently price carry-driven risk premia but leave room for ML to enhance price return forecasts, particularly in high-yield segments. For investors, the results caution against direct adoption of equity-style ML strategies in bonds unless coupled with transaction cost mitigation. Methodologically, we advance the literature by integrating carry decomposition into ML frameworks—a approach that could extend to other fixed-income assets.

Finally, this paper is related to previous studies that focus on the yield-based estimates of expected corporate bond returns in their asset pricing tests (Campello, Chen, and Zhang, 2008; de Jong and Driessen, 2012; Bongaerts, de Jong, and Driessen, 2017). They argue that this forward-looking measure overcomes some well-known limitations associated with *ex post* average returns. Our results offer some empirical justification for their methodology from another perspective: at least for these characteristics having been shown to possess significant in-sample predictive power, they do not capture a larger fraction of forecastable variations in corporate bond returns, in real time, than the simple yield-based estimates.

The remainder of the paper is organized as follows. Section 2 describes the data sets used in our empirical analysis. Section 3 presents our empirical design. Section 4 focuses on the forecasting performance of machine learning models. In Section 5, we assess the economic value of machine learning forecasts. Section 6 concludes.

2 Data

2.1 Data Sources

Since common machine learning practice requires large data sample, we follow the recent studies in the corporate bond literature (Chordia et al., 2017) and combine multiple data sources to extend the sample back to pre-Trace era. Our compiled data set comprises corporate bond returns from the following six data sources:

- The Lehman Brothers fixed income database covers month-end bid prices from January 1973 to March 1998. We follow previous studies (e.g., Eom et al. (2004)) by excluding "matrix" prices, which derived from price quotes of bonds with similar characteristics (Warga and Welch, 1993).

- Beginning in 1994, the National Association of Insurance Commissioners (NAIC) began providing transaction data based on Schedule D filings by all of its member insurance companies. Many studies (e.g., (Hong and Warga, 2000; Campbell and Taksler, 2003)) indicate that these transactions compose a significant portion of the market for publicly traded corporate bonds. We access the NAIC database through Mergent FISD and remove transactions arising from the exercise of call options.

- The BofAML database provides daily quotes for individual bonds covered ICE BofA bond indices, with the sample period spanning from January, 1997. These quotes are commonly used by bond mutual funds and other institutional investors to mark their portfolios to market. Following Schaefer and Strebulaev (2008), we focus on bonds included into either the ICE BofA investment-grade index (MERC0A0) or the ICE BofA US High Yield Index (MERH0A0).

- Starting from July 2002, the Trade Reporting and Compliance Engine (TRACE) provides comprehensive coverage of transactions for publicly traded OTC corporate bonds. We extract data from the enhanced version of TRACE, which includes more HY bond transactions in early years and more precise information about transaction volumes than the "standard" TRACE does. The algorithm of Dick-Nielsen (2013) is employed to identify and correct reporting errors in TRACE data, and we restrict our sample to bond transaction data without duplicates, reversals, and corrections/cancellations.

- Reuters Fixed Income Database collects daily quotes provided by major dealers in the U.S. corporate bond market since 1991. Choi (2013) and Choi and Richardson (2016) show that at monthly frequency Reuters dealer quotes reflect transaction prices quite well. Unlike the Lehman and BofAML data, the Reuters data does not target a specific set of bonds that comprise their indices.

- The IHS Markit bond pricing database collects daily price information from more than thirty dealers to compute a composite price. Similar to the the Reuters database, it is supposed to provide a comprehensive picture of the corporate bond universe and therefore extensively used in studies on the secondary market liquidity (Friewald et al., 2012; Schestag et al., 2016). Our sample of the Markit data spans from January, 2003 to June, 2019.

While the Lehman data are backdated to 1970s, we do not pursue forecasting bond returns over the pre-1991 period, in which daily price observations were generally unavailable. The reason is that the calculation of bond-level illiquidity measures, e.g., the gamma of Bao et al. (2011), requires at least daily price observations, if not transaction-by-transaction information. However, we do make use of the Lehman data before 1990s to construct some bond-level characteristics (e.g., long-term reversal and idiosyncratic volatility), which requires monthly bond returns over the past few years. To ensure that these characteristics have sufficient observations in early years, we focus on the sample between July 1992 and June 2019 for our empirical analysis.

2.2 Individual Bond Returns

From each database, we collect consecutive monthly prices and compute the month-t bond return as

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

where P_t denotes the clean price at the end of month t, AI_t is the corresponding accrued interest, and $C_{i,t}$ is the coupon paid in month t. To compute monthly returns for TRACE and NAIC, we follow the standard practice in the literature by calculating P_t as the volume-weighted average of all trades within five trading days of the month-end (e.g., Chordia et al. (2017))[1] Regarding quote-based databases, the Lehman data is already of the monthly frequency, and we simply take the last available daily price from other databases, given that it take places within the last five trading days of the month. When there are returns for the same bond-month available from multiple sources, we take the first available return in the following sequence: TRACE, NAIC, BofAML, Lehman, Markit, and Reuters. This sequence gives precedence to trade-based returns.

We apply the following standard filters to the combined data set of bond returns: (1) the issue's Mergent bond type is in "US corporate debentures" or "US corporate MTN" categories (CDEB or CMTN, respectively); (2) bonds are denominated in US dollars, senior unsecured, and with a fixed coupon rate; (3) bonds are neither convertible nor exchangeable, and (4) bonds have at least one year to maturity. These filters lead to a union of 44,809 bonds issued by 6,021 companies over the 1992-2019 period. The filtered sample contains an average of 6,895 bonds per month, arguably the largest cross section among recent studies on corporate bond returns.

2.3 Individual Bond Carry and Price Returns

Following Jensen et al. (2018), we decompose corporate bond returns into price returns and yield-implied carry.

[1] It is also equivalent to the approach of Bao and Pan (2013), who adopt the trade-size weighted average of clean prices within the last seven calendar days of the month.

By definition, bond yields are the expected returns of holding the bond to maturity if no credit event occurs. Assuming that the term structure of expected returns is flat,[2] an estimate of expected excess returns can be derived by correcting for the expected default loss,

$$\widetilde{rx}_{i,t+1} = (1 + y_{i,t})\left(1 - (1 - R_t)\pi_{i,t}(T - t)\right)^{1/(T-t)} - (1 - rf_t), \tag{2}$$

where $\pi_{i,t}(T-t)$ is the default probability of bond i with $T-t$ years to maturity, R_t denotes the recovery rate given default, and the risk-free rate rf_t is proxied by the one-month T-bill rate at time t.

There are many ways to obtain estimates of the default probability for individual bonds, including (1) the Merton distance to default (D2D) as well as its variants as examined by Bharath and Shumway (2008), (2) the doubly-stochastic hazard model with observable coparties (Duffie et al., 2007), (3) the Moody's expected default frequency (EDF) score which combines D2D with information from equity markets and financial statements. And most of these estimation methods require the bond issuers to be publicly listed firms. In this paper, we employ a simple estimator of default probabilities which is applicable to all bonds. That is, for yield observations of bonds with letter rating k in year j, we calibrate their default probabilities to the Moody's cumulative issuer-weighted default rate for rating k over the period from 1970 to year $j-1$, with the time to maturity rounded to the nearest integer. While we calculate these cumulative default rates by retrieving data from the Moody's Default & Recovery Database (DRD), they could also be derived from the (annual) cohort-level statistics as included in the appendix of Moody's annual default studies. Therefore, our estimates of default probabilities can be easily derived using public information, without knowledge on the structural or reduced-form models of credit risk.

In a similar manner (yet with lower separating capacity), the recovery rate is assumed the same for all bond observations in year j and calibrated to the Moody's issuer-weighted average recovery rates for senior unsecured bonds up to year $j-1$. Note that the temporal

[2] Bongaerts et al. (2017) empirically investigate the slope of the term structure of expected bond returns and find a slope coefficient of 2.5 basis points per duration-year. As indicated by our unreported results, incorporating this slope coefficient into our construction of yield-implied expected bond returns leads to a slight improvement in their prediction accuracy.

variations in cumulative default rates and recovery rates reflect macroeconomic conditions (Chen, 2010) as well as the updating of beliefs (Benzoni et al., 2015). For example, the calibrated 10-year default probability for AA-rated bonds is about 0.57% for bonds in 2005 and increases to 0.84% for bonds in 2010.

We define the price return as the difference between bond returns in Eq. (1) and yield-implied carry in Eq. (2).

2.4 Bond Return Predictors

Following Lewellen (2015), we focus on characteristics that have been demonstrated by existent studies as significant bond return predictors. Specifically, we survey the credit bond literature for a list of candidate "anomaly" characteristics, both at the security level and at the issuer level. We then narrow the list down to ensure adequate data coverage going back in time. We ultimately settle on 60 bond characteristics and 153 equity (firm-level) characteristics by Jensen et al. (2023). Appendix A provides a detailed discussion and definition of all these bond and stock characteristics.

The first category of predictors is bond characteristics, which covers basic bond information such as size, credit rating, and duration. The effect of credit rating and duration have been exhaustively examined since the first few studies on the cross section of individual bond returns (Gebhardt et al., 2005b,a). Houweling and Van Zundert (2017) show that size-sorted portfolios generate alphas in the corporate bond market. It also covers past bond returns, such as bond momentum (Jostova et al., 2013), downside risk and short-term reversal, long-term reversal (Bali et al., 2021b), and skewness. We also consider security-level illiquidity. To address the concern that the gamma estimated with lower-frequency data might be less informative than its counterpart with transaction data (Dick-Nielsen et al., 2012), we consider alternative liquidity proxies as well. The first one included is Hasbrouck (2009)'s measure based on Gibbs sampling. As shown by Schestag et al. (2016), among tens of low-frequency measures for transaction costs (constructed with daily data) in the corporate bond market, the Hasbrouck measure has the lowest mean bias against many high-frequency benchmarks (constructed with transaction data).

The second category includes 153 firm-level characteristics in Jensen et al. (2023). For example, Chordia et al. (2017) find that asset growth, profitability, equity reversal, and equity momentum posses significant predictive power for corporate bond returns after controlling for bond characteristics. Also, results of Choi and Kim (2018) indicate that asset growth and investment-asset ratio are negatively related to corporate bond returns, and that the effect of equity momentum is positive. Finally, Israel et al. (2018) show that three measures of safety—market leverage, gross profitability, and duration—are priced in the cross section of corporate bond returns.

2.5 Summary Statistics

As shown in Table 1, the combined data sample contains nearly 1.26 million observations of monthly bond returns. The mean and standard deviation of monthly excess returns are 0.40% and 2.60%, largely consistent with the magnitude as reported in previous studies. The reported percentiles show that, unlike equity returns, the distribution of corporate bond returns is rather symmetric, exhibiting merely a small degree of right skewness. The average of monthly price returns is 0.10%, with a standard deviation of 2.50%, while the average of monthly yiled implied carry is 0.40%, with a standard deviation of 0.20%.

[Insert Table 1 Here]

It is worth noting that our compiled data set is largely immune to the sample selection bias arising from the illiquidity of the secondary corporate bond market. Specifically, if only the transaction data is used to study corporate bond returns, a bond could have quite a few missing observations from time to time before it is removed from the sample due to default or some maturity restriction. This issue is not a concern in studies on the long-term performance of corporate bonds, but might give rise to selection bias when testing the cross-sectional predictability. In particular, if transactions occur only when the value of the information signal is sufficient to exceed the trading costs, as modelled in Lesmond et al. (1999) and Chen et al. (2007), then those missing observations essentially reflect an increase in transaction costs, which could result in a large price discount if a transaction has to be made (He and Xiong, 2012; He and Milbradt, 2014). It follows that a portfolio formed at time t only with bonds having time-$t+1$ return observations (this information is unknown at time t) is likely to have an upward bias in its performance assessment. Since our comprehensive data set consists of transactions and dealer quotes from six data sources, most issues have zero missing observations of monthly returns. For comparison, we examine the stability of the bond cross sections in both our data set and the Trace data. Over the same period (from July 2002 to June 2019), about 74% of bonds in our data set always have consecutive monthly returns throughout their presence in our sample, whereas this fraction is merely 24% for the Trace sample.

3 Empirical Methodology

Our primary tests focus on monthly return forecasts derived from machine learning models and from contemporaneous bond yields. In this section, we first summarize the formation of different return forecasts and then examine their effectiveness in out-of-sample tests. Especially, we assess the impact of model choice on the absolute prediction performance as well as the relative performance against the yield-based return forecasts.

3.1 Estimation Strategy

Following Bianchi et al. (2021), we consider a variety of machine learning methods with increasing complexity. For penalized regressions, we implement Ridge regressions, Lasso (Tibshirani, 1996), and Elastic Net (Zou and Hastie, 2005). For linear dimension reduction techniques, we use principal component regressions (PCR) and partial least squares (PLS). To handle nonlinearity and interaction effects, we consider tree-based methods, which include XGBoost, random forests (Breiman, 2001), and ExtraTrees. Appendix B provides a detailed description of these methods.

As standard in the literature, we estimate machine learning models recursively by dividing our data into three disjoint subsamples. We keep the temporal ordering of the data, such that the validation set directly follows the training set and the testing sample directly follows the validation set. Our sample splitting scheme generally resembles previous studies: the initial proportion of in-sample data is roughly equivalent to that of out-of-sample data (12 years v.s. 15 years). We fix the length of validation sample at two years. In other words, our testing sample starts from July 2004.

Since many machine learning algorithms are computationally intensive, we avoid tuning hyperparameters every month. Instead, we assume that the optimal hyperparameters for a model are stable in a short period and re-tune these hyperparameters every year. Each time we re-tune, we increase the training sample by one year, and maintain the same size of the validation sample, but roll it 12 months forward. As most of our bond characteristics are updated monthly, we do refit our model each month using the hyperparameters most recently tuned to incorporate new information.

For example, at the end of 2004:06, we split the sample into 10 years of training sample (1992:07-2002:06), and 2 years of validation sample (2002:07-2004:06) to choose the optimal hyperparameters and make out-of-sample predictions for 2004:07. While at the end of 2004:07, we just adopt the hyperparameters tuned at the end of 2004:06 to refit our model and predict one-month-ahead corporate bond returns. Note that since there is no need to validate the hyperparameters, the whole sample period (2002:07-2004:07) is used as the training sample. This set of hyperparameters is used to refit the model each month until the end of 2005:06, when we validate the hyperparameters using the training sample (1992:07-2003:06) and the validation sample (2003:07-2005:06). This estimation strategy can be viewed as a combination of those used by Gu et al. (2020) and Bianchi et al. (2021), given the trade-off between computational burdens and timely information updating.[3]

[3]Gu et al. (2020) periodically refit the model once per year and make out-of-sample predictions using the same fitted model over the subsequent year, while Bianchi et al. (2021) recursively refit the model every month.

3.2 Performance Evaluation

The predictive performance of machine learning model $\mathcal{M}_k$ for bond i is evaluated by the out-of-sample R-squared

$$R^2_{OOS} = 1 - \frac{\sum(rx_{i,t} - \widehat{rx}_{i,k,t})^2}{\sum rx^2_{i,t}}, \tag{3}$$

where $rx_{i,s}$ denotes the excess return on bond i at time t, and $\widehat{rx}_{i,k,t}$ is the return forecast produced by model $\mathcal{M}_k$. Note that the denominator in Eq. (3) is the sum of squared excess bond returns without demeaning (Gu et al., 2020). R^2_{OOS} measures the reduction in the mean squared forecast error compared to a naive benchmark of zero excess returns for all convertible bonds. A positive R^2_{OOS} indicates that the machine learning forecast model outperforms the naive benchmark.

A natural benchmark for the evaluation of regression-based predictions is the yield-implied expected bond returns (Campello et al., 2008; de Jong and Driessen, 2012; Bongaerts et al., 2017; van Zundert and Driessen, 2022).[4] We repeat the exercise of return prediction by replacing $\widehat{rx}$ with $\widetilde{rx}$ in Eq. (2) and report the estimated R^2_{OOS}.

Next, to obtain statistical inference, we compare the forecasts of two models in terms of out-of-sample performance, following Diebold and Mariano (1995). The test statistic is defined as:

$$DM^{12} = \frac{\overline{d}^{12}}{\hat{\sigma}^{12}_{\overline{d}}}, \tag{4}$$

where $\overline{d}^{12}$ and $\hat{\sigma}^{12}_{\overline{d}}$ denote the time-series average and Newey and West (1987) standard errors of the mean difference between squared forecast errors d^{12}:

$$d^{12}_{t+1} = \frac{1}{n_{3,t+1}} \sum_{t=1}^{n_{3,t+1}} ((\hat{e}^{(1)}_{i,t+1})^2 - (\hat{e}^{(2)}_{i,t+1})^2), \tag{5}$$

where $\hat{e}^{(1)}_{i,t+1}$ and $\hat{e}^{(2)}_{i,t+1}$ denote the prediction errors for bond i at time t using models 1 and 2, respectively, and $n_{3,t+1}$ is the number of bonds in the testing sample.

[4] Bongaerts et al. (2011) employ a similar method to derive an estimate of expected excess returns on CDS portfolios.

4 Bond Return Predictability

In this section, we examine the predictability of corporate bond returns with various ML methods and focus on the relative performance of the carry-augmented approach. Sections 4.1 and 4.2 present the forecasting accuracy in samples of listed and unlisted bond issuers, respectively. Section 4.3 considers another dimension and reports the subsample analysis on investment-grade and high-yield bonds. Section 4.4 discuss the improvement of the carry-augmented approach from an econometrical point of view. Section 4.5 assesses variable importance in each machine learning algorithm.

4.1 Bonds Issued by Listed Firms

Panel A of Table 2 presents the monthly R^2_{OOS} in percentages using the 60 bond characteristics and 153 stock characteristics as covariates and covering the entire sample of public corporate bonds. Results in the first row show the bond excess return predictability. We find that all machine learning algorithms show some degree of predictive power for individual bond returns, with the R^2_{OOS} ranging from 1.16% (PCR) to 2.24% (Enet). The forecast combination model (Comb), which averages different types of machine learning forecasts, also performs well and results in an R^2_{OOS} value of 2.29%. The standard linear regression estimated via ordinary least squares (OLS) deliver disastrous performance, with the R^2_{OOS} of 2.14%. This result echoes Lewellen (2015)'s finding from the stock market: return predictors that have been screened by the empirical literature likely to contain independent determinants of expected returns, which have substantial predictive power even in the OLS setting.

[Insert Table 2 Here]

However, the last column shows that the superior performance of machine learning techniques is eclipsed by a simple benchmark—yield-implied expected bond returns. Indeed, the R^2_{OOS} generated by $\widetilde{rx}$ (2.95%) is much greater than that of the best-performing machine learning model. Note that our yield-based estimate does not contain the "roll down" component, so it is even expected to underpredict the realized bond returns.

Next, we decompose bond total returns into carry and price returns, and predict price returns using machine learning methods. We compare the sum of carry parts and predicted price returns with bond total returns. The results in the second row in Panel A reveal that the predictive performance is much better than that with bond returns for each machine learning algorithm. The R^2_{OOS} of all machine learning methods ranges from 2.73% (PCR) to 3.68% (Enet), which is 1.26-1.53% higher than that for predicting for corporate bond returns in the first row. The forecast combination model (Comb) also generates an R^2_{OOS} value of 3.63%. Almost all methods have better predictive performance than the simple benchmark—yield-implied expected bond returns (R^2_{OOS}=2.95%).

4.2 Bonds Issued by Unlisted Firms

In Panel B of Table 2, we only consider the 60 bond characteristics into machine learning models. Accordingly, the sample is confined to private bonds. We find that excluding stock characteristics to some extent downgrades the out-of-sample performance. For example, for predicting bond excess returns, the R^2_{OOS} of Enet decreases from 2.24% to 1.78%, and that of Combination decreases from 2.29% to 2.05%. The yield-based estimates generate an R^2_{OOS} of 3.61% over the entire private firm sample, which outperforms machine learning methods.

In terms of predicting prices returns, we find that the predictive performance is much superior than that with bond returns for each machine learning algorithm. The R^2_{OOS} of all machine learning methods ranges from 3.21% (XGBoost) to 3.54% (OLS), which is 1.23-1.81% higher than that for predicting for corporate bond returns. It is comparable to the out-of-sample performance of price return prediction in public firms. The forecast combination model (Comb) also generates an R^2_{OOS} value of 3.48%. All methods have inferior predictive performance than the simple benchmark—yield-implied expected bond returns, with an R^2_{OOS} value of 3.61%.

In summary, when predicting prices returns among both public and private firms, we find that all machine learning algorithms have better predictive performance than that for predicting bond total returns. The simple benchmark—yield-implied expected bond returns yields better predictive performance than machine learning models except for predicting price returns among public firms.

4.3 Investment-Grade v.s. High-Yield Bonds

In addition to the full sample of bonds, we examine the predictive performance for investment-grade and high-yield bonds separately, and report the corresponding results in Table 3. We obtain similar patterns in R^2_{OOS} for investment-grade and high-yield bonds for all machine learning models. For investment-grade bonds in Panel A, when benchmarked against the yield-implied expected returns to predict bond total returns, all machine learning methods generally underperform. For predicting price returns, the yield-implied expected returns achieve the highest out-of-sample R^2 for private firms, compared with all machine learnings methods.

[Insert Table 3 Here]

Regarding high-yield bonds in Panel B, penalized regression models and the forecast combination model achieve higher R^2_{OOS} values compared with other machine learning methods. Their predictive performance is inferior to that of the benchmark for predicting bond excess returns, but superior than the benchmark—yield-implied expected bond returns for predicting bond price returns.

Overall, we find that machine learning is dominated by the yield-based benchmark when predicting excess returns in both the investment-grade and high-yield bond subsamples, but shows superiority in forecasting price returns.

4.4 Forecast Comparison

In addition to quantitatively comparing the predictive performance of machine learning models in Table 2, we adopt the Diebold and Mariano (1995) tests to iteratively compare the forecasts of two competing models, following Eq. (4). Under the null hypothesis that there is no difference between the two models, the Diebold-Mariano statistics follow a N(0,1) distribution. The magnitudes of the test statistics correspond to p-values in the same way as t-statistics in regression analysis.

We present the results in Table 4, which reveal two interesting findings. First, in the first row, we compare whether price return prediction based on all machine learning methods outperforms the excess return prediction with all machine learning methods. The positive Diebold-Mariano statistics reveal that price return prediction outperforms the excess return prediction with all machine learning methods. The statistics are significant for private firms in Panel B. Second, in the second row, we compare whether yield-implied expected returns outperform machine learning predicted returns when predicting excess returns. We find that yield-implied expected returns outperform machine learning predicted returns. The statistics are significant for private firms in Panel B.

[Insert Table 4 Here]

4.5 Which Predictors Matter?

Given the large number of predictors, we next investigate which predictors are more important. We adopt the reduction in panel predictive R^2 to gauge variable importance, following Gu et al. (2020). Figures 1 and 2 illustrate the overall importance of all predictors based on the pooled full sample. We calculate the reduction in R^2 from setting all values of a given predictor to zero within each training sample and average these into a single importance measure for each predictor. We order predictors along the vertical axis by calculating the sum of the ranks of R^2-based variable importance across different models and sorting them from highest to lowest. This ordering reflects the overall contribution of a predictor to all models. Each column corresponds to a prediction model. The color gradient within each column indicates the model-specific ranking of predictors from most to least important (darkest to lightest).

[Insert Figures 1 and 2 Here]

Regarding overall variable importance ranks, machine learning models indentify short term reversal, relative value, and downside risk as key bond characteristics, when predicting excess returns of public bonds. Equity characteristics, such as asset turnover, book to market and firm leverage are also important predictors, which indicates the integration between the stock and bond markets. Similar patterns are observed for predicting price returns. In the case of private bonds, we find that reversals, duration, downside risk and illiquidity are the most important predictors.

5 Economic Values

Our assessment of prediction performance in Section 4 is entirely statistical. This raises the question of whether the positive out-of-sample R^2 statistics is economically meaningful for investors. This section addresses this questions and, especially, examines if machine learning forecasts become superior to yield-based ones once we consider economic gains. In Section 5.1, we form bottom-up forecasts for prespecified portfolios and evaluate their economic values with a market timing trading strategy. Section 5.2 focuses on long-short portfolios based on different types of return forecasts and highlights the impact of trading costs on their actual performance.

5.1 Asset Allocation with Prespecified Portfolios

Compared to individual bond issues, bond portfolios is likely to be of broader economic interest as they represent the risky-asset savings vehicles most commonly held by investors in the corporate bond market. Therefore, we build bottom-up return forecasts by aggregating machine learning forecasts into prespecified portfolios. We consider the 12 double-sorted, monthly-rebalanced portfolios. Specifically, individual bonds are firstly grouped into 16 rating categories: AA+ (including AAA and AA), A, BBB, and speculative grades (BB and below). And bonds in each rating category are further assigned one of three maturity groups: short($<$ 5 years), intermediate (5-10 years), and long ($>$ 10 years). Portfolio returns are value-weighted average returns of bonds in each portfolio. These rating-maturity portfolios are deeply rooted in the real corporate bond market, as many bond market index families design their subindices in this way. For instance, ICE BofA fixed income indices include ones like “10+ Year BBB US Corporate Index”, and Barclays has a similar one termed “Long Baa US Corporate Total Return Index”. Similarly, many bond mutual funds and ETFs have restrictions on their investment from a rating and/or duration perspective (e.g., PIA BBB Bond Fund and Vanguard Intermediate Term Corporate Bond ETF).

Before proceeding to the economic gains associated with various portfolio return forecasts, we examine their accuracy as it could markedly differ from the results of security-level forecasting. As discussed in Gu et al. (2020), the distribution of portfolio returns is sensitive to dependence among individual security returns. As such, some components of security-level forecasting errors tend to be cancelled out at the portfolio level, while other components might be amplified.

Results in Figures 3 and 4 indicate that machine learning methods generate positive out-of-sample R^2s for most rating-maturity portfolios among public and private firms, respectively. When predicting bond excess returns among public firms in Panel A of Figure 3. Machine learning methods are outperformed by the yield-implied expected bond returns for high-yield bonds, especially for short-term bonds. As discussed in Section 4, corporate bonds with lower credit ratings are conceptually closer to defaultable bonds, and short maturities make the bias of yield-implied expected returns less severe. On the other hand, the advantage of machine learning emerges in high-yield grades, which is also consistent with the findings of security-level forecasts. When predicting price returns among public firms in Panel B, both machine learning and the yield-implied expected bond returns generate the highest out-of-sample R^2s for high-yield bonds, especially for short-term bonds.

[Insert Figures 3 and 4 Here]

For private firms in Figure 4, we find that the yield-implied expected bond returns for A-rated and high-yield bonds, especially for short-term bonds, outperform the machine learning methods, when predicting excess returns. The machine learning methods and the benchmark are comparable and generate the best performance for A-rated and high-yield bonds, especially for short-term bonds among private firms

Compared to the statistical significance of portfolio return predictability, we are more interested in its economic values. To this end, we follow Campbell and Thompson (2008) by evaluating trading strategies that time each prespecified portfolio with machine learning forecasts. Specifically, we consider a mean-variance investor who selects a particular rating-maturity portfolio along with a risk-free asset. It follows that the optimal weight on the prespecified portfolio with machine learning method $\mathcal{M}_k$ is given by

$$w_{k,t} = \frac{E_t^{\mathcal{M}_k}(xr_{t+1})}{\gamma E_t \left(xr_{t+1} - E_t^{\mathcal{M}_k}(xr_{t+1}) \right)^2}, \tag{6}$$

where γ denotes the coefficient of relative risk aversion. To proxy for the conditional variance in Eq. (6), we employ a rolling sample variance estimator as in Thornton and Valente (2012). With the estimated optimal weight on the risky bond portfolio, we compute the realized utility for $\mathcal{M}_k$ and benchmark it against the yield-based return forecasts.

To test whether the welfare gains brought about by $\mathcal{M}_k$ are significantly greater than the benchmark, we conduct a variant of Diebold and Mariano (1995) test as proposed by Bianchi et al. (2021).[5] To be more specific, we calculate the certainty equivalent return (CER) values for each month in the out-of-sample period and then estimate the following regression

$$u_{k,t} - u_{Y,t} = \nu + \varepsilon_t, \tag{7}$$

where $u_{k,t}$ and $u_{Y,t}$ represent the realized utilities from investing with $\mathcal{M}_k$-based and yield-based return forecasts, respectively. This asset allocation exercise is rather similar to out-of-sample tests of equal forecast accuracy and encompassing, except that the loss function is not directly based on forecast errors but on the mean-variance utility.

Figures 5 and 6 present the CER values (in percentage points) with γ set to 5, a typical value adopted by asset allocation studies in the fixed income field (Koijen et al., 2010; Thornton and Valente, 2012). We also follow these studies by imposing short-sales and borrowing constraints. Relative to the yield-implied benchmark, trading on machine learning forecasts of bond excess return does not generate significant utility gains for the mean-variance investor. However, trading on machine learning forecasts of bond price return plus

[5] Bianchi et al. (2021) propose using the test of Harvey et al. (1997) to account for autocorrelation in the forecasting errors.

carry does improve the investors' utility, especially for high-yield bonds, where the CER values are positive and significant at the 5% level for most of machine learning methods.

[Insert Figures 5 and 6 Here]

This finding is robust if we exclude equity characteristics and confine the sample to private firms. For most of machine learning models with price return as target, they perform superior to the yield-implied expected bond returns, especially for intermediate and short term high yield bonds.

5.2 Hedged Portfolio Returns and Trading Costs

Besides prespecified bond portfolios, we also construct long-short portfolios based on machine learning forecasts to assess their economic values from an alternative perspective. For each machine learning method, we sort corporate bonds into deciles based on its return predictions and then form zero-net-investment portfolios that are long in the highest expected-return bonds and short in the lowest ones. Table 5 displays the profitability of these hedged portfolios.

[Insert Table 5 Here]

Panel A reports the results using all bond and stock characteristics in public firms. When predicting bond excess returns, we find that penalized linear regression models deliver the best performance among machine learning models, generating a spread in average monthly returns of 0.82%-1.11% between the top and bottom deciles. Close to these best-performers is the mean combination of individual model forecasts, with the average monthly return up to 0.97%. All of the machine learning models lead to less gross profitability than the yield-based benchmark. Panel B reports the results using all bond characteristics in private firms, which display similar findings.

The superb performance of the yield-based benchmark portfolios is further demonstrated by the risk-adjusted returns. That is, we estimate alphas from the time-series regressions of portfolio returns on the corporate bond market factors. The alpha for penalized linear regression models and the combination model ranges from 0.65% to 1% with both bond and stock characteristics and from 0.52% to 0.65% with only bond characteristics. Therefore, standard risk factors cannot explain the gross investment gains from machine learning portfolios, which is consistent with the findings of Feng et al. (2025). On the other hand, the gross return on the benchmark portfolio is not absorbed by its exposures to common risk factors. While the alpha is still significant, its magnitude shrinks to 0.85% per month with both bond and stock characteristics and 0.76% per month with only bond characteristics. This finding points to the efficiency of the corporate bond market, as the current yield seems to already reflect compensations for various sources of risk.

Regarding predicting price returns, we find that penalized linear regression models deliver the best performance among machine learning models. With both bond and stock characteristics among public firms, the portfolio performance ranges between 1.27% and 1.74% per month. It is higher than that for predicting excess returns and than the yield-based benchmark portfolio (1.38%). With only bond characteristics among private firms in Panel B, the portfolio performance ranges between 1.03% and 1.32% per month, which is comparable to the yield-based benchmark portfolio (1.22%). We have similar findings for risk-adjusted returns.

The distinct impact of risk adjustments on hedge portfolio performance can also inferred from cumulative portfolio returns presented in Panels A1 and B1 of Figures 7 and 8.[6] The machine learning portfolio and the benchmark portfolio generate comparable cumulative returns over the 15-year period, whereas the latter suffers much larger losses during the 2008 financial crisis. It suggests that that bond yield already reflects compensations for various sources of risk—market risk, default risk, downside risk, and liquidity risk—which are particularly relevant in financial crisis. While our yield-based benchmark is adjusted for expected default loss, there is a well-known, sizable wedge between physical and risk-neutral expectations of default loss—which lies at the heart of the credit spread puzzle (Huang and Huang, 2003)—and a large fraction of this wedge is associated with economic recessions (Chen et al., 2009; Chen, 2010; Bhamra et al., 2010). As such, an adjustment for physical default loss is clearly insufficient to remove the compensation for macroeconomic risks.

[6]For simplicity, we focus on the machine learning portfolio implied by forecast combination, as a balanced choice between the gross return and (as will be discussed below) the return net of transactions costs.

[Insert Figures 7 and 8 Here]

In the corporate bond market, trading costs are far from negligible when translating theoretical portfolio returns to real profitability. For this reason, we also report hedge portfolio returns net of transaction costs in Table 5. Following Chordia et al. (2017), we compute the portfolio-level cost as the product of the turnover rate and some measure of effective (fractional) bid-ask spread. As shown by previous works (e.g., Bao et al. (2011) and (Dick-Nielsen et al., 2012)), there is substantial temporal variations in bond-level transaction costs. To undertake a more accurate assessment of the liquidity effect, we first use the bid-ask spread of individual bonds to calculate the average transaction costs for each rating portfolios month by month, and use the costs to proxy for transaction costs of a bond with corresponding rating.

$$\lambda_t = \frac{\xi_t}{N_t} \Sigma_{i=1}^{N_t} k_{i,t},$$

where ξ_t denotes the month-t portfolio turnover, $k_{i,t}$ the security-level transaction cost, and N_t the number of bonds in the portfolio in month t. The bid-ask spread is estimated by Gibbs sampling of daily quotes following Hasbrouck (2009), which measures bond trading cost well according to Schestag et al. (2016).

Results for predicting bond excess returns in column (3) of Panel A indicate that trading on machine learning forecasts involves high portfolio turnover, ranging from 0.77 to 1.43. The resultant trading costs overwhelm the significant gross returns on these hedge portfolios. Seven of 9 machine learning portfolios have significantly positive returns net of trading costs, with the raw returns ranging between 0 and 0.41% per month. The benchmark portfolio outperforms all machine learning portfolios by a large margin, with the net return of 1.15% and a turnover of 0.42.

The portfolio turnover decreases when equity characteristics are excluded into machine learning models, as shown in Panel B. It results in a lower gap between the gross and net returns. Indeed, seven machine learning portfolios maintain positive profitability, when the trading cost is estimated using the Hasbrouck model. Consistent with the evidence in Panel A, the composition of the benchmark portfolio is fairly stable over time, with the average turnover of 0.42. Consequently, its net returns (1.15% per month) overwhelmingly predominate among all long-short portfolios.

When predicting price returns, we obtain similar findings. After considering transaction costs, all machine learning portfolios become lower and remain positive returns, ranging between 0.49% and 1.10% per month for both public and private firms. However, they underperform the benchmark portfolio with the net return of 1.15% per month.

The impact of transaction costs on the investment gains is visualized by Panels A2 and B2 of Figures 7 and 8, in which relative performance of machine learning portfolios and the benchmark portfolio is overturned.

The evidence so far indicates that it is largely infeasible to exploit the profitability associated with these zero-net-investment portfolios, which is seriously undermined either by risk adjustments (for the benchmark portfolios) or by market illiquidity (for machine learning portfolios). We further combine these economic forces by examining if these portfolios generate significantly positive alphas net of trading costs. With the Hasbrouck measure adjustment, only half of the machine learning hedge portfolios and yield-implied expected returns yield significantly positive alphas. And both lose the significance when the Hasbrouck model is used to estimate transaction costs. Results among both private and public firms for both predicting excess returns and price returns carry the same message. Overall, our findings seem to support the notion that the corporate bond market is generally efficient net of transaction costs.

6 Conclusion

The asset-pricing literature finds significant cross-sectional predictability in corporate bond returns. This paper provides new evidence on their real-time predictability as uncovered by machine learning models. With a comprehensive sample of more than 1.3 million U.S. bond-month observations from 1992 to 2019, this study compares the out-of-sample predictive power of machine learning algorithms and the yield implied expected returns based on 60 bond characteristics and 153 equity characteristics. It also decomposes bond excess returns into carry parts and price returns, and predicts price returns and bond excess returns separately.

It documents that the out-of-sample performance for all machine learning methods for predicting price returns is much better than that for predicting bond excess returns. However, we also find that the expected return estimates as derived from the current yield to maturity appear to outperform machine learning characteristics-based forecasts, especially for predicting bond excess returns.

Moreover, we construct long-short portfolios based on predicted returns. Only the yield-based estimates of bond premium offer robust investment gains after accounting for exposures to systematic risks and transaction costs. Our results suggest that the corporate bond predictability is largely reconcilable with exposure to standard risk factors and friction-induced mispricing.

For a practical perspective, our findings call for direct optimization of corporate bond portfolios in practice. That is, the supervised learning approach as examined in this paper essentially performs indirect optimization, which entails estimation of asset risk premiums in the first step. We find that an accurate inference of expected bond returns does not necessarily lead to sizable economic gains, highlighting the importance of directly learning portfolio weights and allowing for flexible learning objectives (rather than prediction errors). More importantly, the documented massive impact of transaction costs on investment points to the necessity of taking the trading environment into account. Given these insights, reinforcement learning as arising in the literature appears a perfect recipe to fill these gaps (Cong et al., 2021). We leave for future studies the potential improvement brought by reinforcement learning in bond investment performance.

References

Bali, Turan G., Heiner Beckmeyer, Mathis Moerke, and Florian Weigert, 2023, Option return predictability with machine learning and big data, *Review of Financial Studies* 36, 2548–3602.

Bali, Turan G., Amit Goyal, Dashan Huang, Fuwei Jiang, and Quan Wen, 2021a, The cross-sectional pricing of corporate bonds using big data and machine learning, *Working Paper, Georgetown University* .

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

Bao, Jack, and Jun Pan, 2013, Bond illiquidity and excess volatility, *Review of Financial Studies* 26, 3068–3103.

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

Benzoni, Luca, Pierre Collin-Dufresne, Robert S. Goldstein, and Jean Helwege, 2015, Modeling Credit Contagion via the Updating of Fragile Beliefs, *Review of Financial Studies* 28, 1960–2008.

Bhamra, Harjoat S, Lars-Alexander Kuehn, and Ilya A Strebulaev, 2010, The levered equity risk premium and credit spreads: A unified framework, *Review of Financial Studies* 23, 645–703.

Bharath, Sreedhar T, and Tyler Shumway, 2008, Forecasting default with the merton distance to default model, *Review of Financial Studies* 21, 1339–1369.

Bianchi, Daniele, Matthias Bchner, and Andrea Tamoni, 2021, Bond risk premiums with machine learning, *Review of Financial Studies* 34, 1046–1089.

Bongaerts, Dion, Frank de Jong, and Joost Driessen, 2011, Derivative pricing with liquidity risk: Theory and evidence from the credit default swap market, *Journal of Finance* 66, 203–240.

Bongaerts, Dion, Frank de Jong, and Joost Driessen, 2017, An asset pricing approach to liquidity effects in corporate bond markets, *Review of Financial Studies* 30, 1229–1269.

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

Campbell, John Y., and Glen B. Taksler, 2003, Equity volatility and corporate bond yields, *Journal of Finance* 58, 2321–2349.

Campbell, John Y, and Samuel B Thompson, 2008, Predicting excess stock returns out of sample: can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Campello, Murillo, Long Chen, and Lu Zhang, 2008, Expected returns, yield spreads, and asset pricing tests. ssrn working paper, *Review of Financial Studies* 21, 1297–1338.

Chen, Hui, 2010, Macroeconomic conditions and the puzzles of credit spreads and capital structure, *Journal of Finance* 65, 2171–2212.

Chen, Hui, Zhiyao Chen, and Jun Li, 2025, The debt-equity spread, *Journal of Finance* forthcoming.

Chen, Long, Pierre Collin-Dufresne, and Robert S. Goldstein, 2009, On the relation between the credit spread puzzle and the equity premium puzzle, *Review of Financial Studies* 22, 3367–3409.

Chen, Long, David A. Lesmond, and Jason Wei, 2007, Corporate yield spreads and bond liquidity, *Journal of Finance* 62, 119–149.

Chen, Luyang, Markus Pelger, and Jason Zhu, 2023, Deep learning in asset pricing, *Management Science* 70, 714–750.

Choi, Jaewon, 2013, What drives the value premium?: The role of asset risk and leverage, *Review of Financial Studies* 26, 2845–2875.

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

Choi, Jaewon, and Matthew Richardson, 2016, The volatility of firm's assets and the leverage effect, *Journal of Financial Economics* 121, 254–277.

Chordia, Tarun, Amit Goyal, Yoshio Nozawa, Avanidhar Subrahmanyam, and Qing Tong, 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, Kee H., Junbo Wang, and Chunchi Wu, 2019, Volatility and the cross-section of corporate bond returns, *Journal of Financial Economics* 133, 397–417.

Cong, Lin William, Ke Tang, Jingyuan Wang, and Yang Zhang, 2021, Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai, *Working paper, Cornell University* .

Correia, Maria, Scott Richardson, and Irem Tuna, 2012, Value investing in credit markets, *Review of Accounting Studies* 17, 572–609.

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

de Jong, Frank, and Joost Driessen, 2012, Liquidity risk premia in corporate bond markets, *Quarterly Journal of Finance* 2, 1250006.

Dick-Nielsen, Jens, 2013, How to clean enhanced trace data Working paper, Copenhagen Business School.

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

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 13, 253–263.

Duffie, Darrell, Leandro Saita, and Ke Wang, 2007, Multi-period corporate default prediction with stochastic covariates, *Journal of Financial Economics* 83, 635–665.

Eom, Young Ho, Jean Helwege, and Jing-Zhi Huang, 2004, Structural models of corporate bond pricing: An empirical analysis, *Review of Financial Studies* 17, 499–544.

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

Feng, Guanhao, Xin He, Yanchu Wang, and Chunchi Wu, 2025, Predicting individual corporate bond returns, *Journal of Banking and Finance* 171, 107372.

Filippou, Ilias, David Rapach, David Rapach, Mark P. Taylor, and Guofu Zhou, 2021, Exchange rate prediction with machine learning and a smart carry portfolio, *Working Paper, Washington University in St. Louis* .

Friewald, Nils, Rainer Jankowitsch, and Marti G. Subrahmanyam, 2012, Illiquidity or credit deterioration: A study of liquidity in the us corporate bond market during financial crises, *Journal of Financial Economics* 105, 18–36.

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

Gebhardt, William R., Soeren Hvidkjaer, and Bhaskaran Swaminathan, 2005b, Stock and bond market interaction: Does momentum spill over?, *Journal of Financial Economics* 75, 651–690.

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

Goyenko, Ruslan, and Chengyu Zhang, 2021, The joint cross section of option and stock returns predictability with big data and machine learning, *Working Paper, McGill University* .

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

Harvey, David, Stephen Leybourne, and Paul Newbold, 1997, Testing the equality of prediction mean squared errors, *International Journal of Forecasting* 13, 281–291.

Hasbrouck, Joel, 2009, Trading costs and returns for u.s. equities: Estimating effective costs from daily data, *Journal of Finance* 64, 1445–1477.

He, Zhiguo, and Konstantin Milbradt, 2014, Endogenous liquidity and defaultable bonds, *Econometrica* 82, 1443–1508.

He, Zhiguo, and Wei Xiong, 2012, Rollover risk and credit risk, *Journal of Finance* 67, 391–430.

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

Houweling, Patrick, and Jeroen Van Zundert, 2017, Factor investing in the corporate bond market, *Financial Analysts Journal* 73, 100–115.

Huang, JingZhi, and Ming Huang, 2003, How much of the corporate-treasury yield spread is due to credit risk?, *Working paper, Pennsylvania State University* .

Ilmanen, Antti, and Rafey Sayood, 2002, Quantitative forecasting models and active diversification for international bonds, *Journal of Fixed Income* 12, 40–51.

Israel, Ronen, Diogo Palhares, and Scott A Richardson, 2018, Common factors in corporate bond returns, *Journal of Investment Management* 16, 17–46.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2018, Carry, *Journal of Financial Economics* 127, 197–225.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *Journal of Finance* 78, 2465–2518.

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

Koijen, Ralph S. J., Theo E. Nijman, and Bas J. M. Werker, 2010, When Can Life Cycle Investors Benefit from Time-Varying Bond Risk Premia?, *Review of Financial Studies* 23, 741–780.

Koijen, Ralph S.J., Tobias J. Moskowitz, Lasse Heje Pedersen, and Evert B. Vrugt, 2018, Carry, *Journal of Financial Economics* 127, 197 – 225.

Leippold, Markus, Qian Wang, and Wenyu Zhou, 2021, Machine learning in the Chinese stock market, *Journal of Financial Economics* .

Lesmond, David A., Joseph P. Ogden, and Charles A. Trzcinka, 1999, A new estimate of transaction costs, *Review of Financial Studies* 12, 113–1141.

Lewellen, Jonathan, 2015, The cross-section of expected stock returns, *Critical Finance Review* 4, 1–44.

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

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

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

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

Thornton, Daniel L., and Giorgio Valente, 2012, Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective, *Review of Financial Studies* 25, 3141–3168.

Tibshirani, R., 1996, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society Series B* 58, 267–288.

van Zundert, Jeroen, and Joost Driessen, 2022, Stocks versus corporate bonds: A cross-sectional puzzle, *Journal of Banking and Finance* 137, 106447.

Warga, Arthur, and Ivo Welch, 1993, Bondholder Losses in Leveraged Buyouts, *Review of Financial Studies* 6, 959–982.

Zou, Hui, and Trevor Hastie, 2005, Regularization and variable selection via the elastic net, *Journal of the royal statistical society: series B (statistical methodology)* 67, 301–320.

Appendix A Bond Return Predictors

A.1 Corporate Bond Characteristics

For each issue in our sample, we consider 60 corporate characteristics as return predictors, whose availability largely depends on the liquidity and age of the issue, as well as the status of the issuer.

Table 1: Corporate Bond Characteristics

1. Credit rating	31. CHL
2. Duration, Gebhardt et al. (2005)	32. Stdev of CHL
3. Age, Israel et al. (2018)	33. Schultz High-low measure, Schultz (2000)
4. Maturity	34. BDJ
5. Bond size, Houweling and Van Zundert (2017)	35. EHP effective spread, Edwards et al. (2007)
6. 5% VaR, Bai et al. (2019)	36. ZTD
7. 10% VaR, Bai et al. (2019)	37. FHT liquidity measure, Fong et al. (2017)
8. 5% expected shortfall	38. Gibbs
9. 10% expected shortfall	39. Amihud LF, Amihud (2002)
10. Gamma LF, Bao et al. (2011)	40. Amihud Roll
11. Roll LF, Roll (1984)	41. Amihud Gibbs
12. Gamma HF, Bao et al. (2011)	42. Amihud Effective tick
13. Roll HF, Dick-Nielsen et al. (2012)	43. Amihud FHT
14. EffectiveTick HF	44. Amihud HLspread
15. Std EffectiveTick	45. Amihud CHL
16. EffectiveTick LF	46. Bond return variance, Bali et al. (2016)
17. Average bid-ask, Hong and Warga (2000)	47. Bond skewness
18. Stdev of average bid ask	48. Bond coskewness, Harvey and Siddique (2000)
19. Interquartile range (IQR), Han and Zhou (2007)	49. Bond idiosyncratic skewness, Boyer et al. (2010)
20. Stdev of IQR	50. Bond kurtosis
21. Round-trip transaction costs (RTC), Feldhutter (2012)	51. VIX beta, Chung et al. (2019)
22. Stdev of RTC	52. Uncertainty beta, Bali et al. (2021)
23. Amihud HF	53. Bond market beta
24. Stdev of Amihud measure	54. Default beta, Chung et al. (2019)
25. Pastor and Stambaugh's liquidity measure, Pastor and Stambaugh (2003)	55. Term beta, Chung et al. (2019)
26. Hasbrouck Lambda, Hasbrouck (2009)	56. 6-month momentum, Jostova et al. (2013)
27. Dispersion	57. 11-month momentum
28. Stdev of Dispersion	58. Long-term reversal, Bali et al. (2021b)
29. High-Low spread, Corwin and Schultz (2012)	59. Short-term reversal, Bai et al. (2019)
30. Stdev of High-Low spread	60. Relative value

This table lists the bond characteristics.

1. **Credit rating:** We retrieve the information of credit rating from Mergent FISD. If a issue has credit ratings from multiple agencies at any point in time, we take the first available rating in the following sequence: S&P, Moody's, and Fitch (Dick-Nielsen

et al., 2012). We convert categorical ratings into conventional numerical scores, where 1 refers to an AAA rating and 2 refers to a AA+ rating. Higher numerical score means higher credit risk. Numerical ratings of 10 or below (BBB- or better) are considered investment grade, and ratings of 11 or higher (BB + or worse) are labeled high yield.

2. **Duration:** The Macaulay duration is a popular measure of term or maturity risk of a bond. We use the modified duration, namely the Macaulay duration divided by one plus the yield to maturity, as our proxy of term risk (Gebhardt et al., 2005a).

3. **Age:** Bond age since the first issuance, in the number of years.

4. **Maturity:** The number of years to maturity.

5. **Bond size:** In the corporate bond literature, Size is typically measured with a bond issuer's total public debt instead of the outstanding amount of individual bonds. Houweling and Van Zundert (2017) find that bonds issued by small companies carry a positive premium. We define a bond issuer's Size in a given month as the sum of the market value weights of all its bonds in our sample.

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

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

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

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

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

10. **Gamma LF:** We follow Bao et al. (2011) to construct the measure, which aims to extract the transitory component from bond price. Specifically, let $\Delta p_{itd} = p_{itd} - p_{itd-1}$ be the log price change for bond i on day d of month t. Then, γ is defined as

$$\gamma = -\text{Cov}_t(\Delta p_{id}, \Delta p_{i,d+1}) \tag{8}$$

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

$$\text{Roll} = \begin{cases} 0 & \text{otherwise,} \\ 2\sqrt{-\text{cov}(r_{i,d}, r_{i,d-1})} & \text{if } \text{cov}(r_{i,d}, r_{i,d-1}) < 0 \end{cases} \tag{9}$$

where $r_d = P_d/P_{d-1} - 1$ denotes the total return on day d.

12. Gamma HF: The intraday version of Gamma.

13. Roll HF: The intraday version of Roll measure of illiquidity.

14. Effective Tick HF

15. Stdev effective Tick

16. Effective Tick LF

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

$$\text{AvgBidAsk} = \frac{\overline{P_t^{\text{Buy}}} - \overline{P_t^{\text{Sell}}}}{0.5 \cdot (\overline{P_t^{\text{Buy}}} + \overline{P_t^{\text{Sell}}})} \tag{10}$$

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

18. Stdev of average buy and sell

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

$$\text{IQR} = \frac{P_t^{75th} - P_t^{25th}}{\bar{P}_t}. \tag{11}$$

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

20. Stdev IQR

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

22. Stdev RTC

23. Amihud HF: The intraday version of Amihud illiquidity measure.

24. Stdev of Amihud: The standard deviation of the daily Amihud measure within a month.

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

$$r_{t+1}^e = \theta + \Theta r_t + \gamma sign(r_t^e) Q_t + e_t, \tag{12}$$

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

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

$$r_\tau = \lambda \cdot \text{sign}(Q_\tau) \cdot \sqrt{|Q_\tau|} + \epsilon_\tau \tag{13}$$

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

$$r_\tau = \alpha \cdot D_i + \lambda \cdot \sqrt{|Q_\tau|} + \epsilon_\tau \tag{14}$$

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

$$D_i = \begin{cases} 1 & \text{if trade } i \text{ is a buy,} \\ 0 & \text{if trade } i \text{ is an interdealer trade,} \\ -1 & \text{if trade } i \text{ is a sell} \end{cases} \tag{15}$$

27. Dispersion

28. Stdev dsipersion

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

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

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

30. Stdev of High-Low spread:

31. CHL

32. Stdev of CHL

33. **Schultz High-low measure:** Schultz (2000) proposes the liquidity measure.

$$\text{Schultz} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(H_i - L_i)^2}}{\frac{1}{n}\sum_{i=1}^{n}(P_i)} \tag{16}$$

34. BDJ

35. **EHP effective spread:** Edwards et al. (2007)

$$\text{Effective spread} = 2 * \frac{|Tradeprice - Midpointprice|}{Midpointprice} \tag{17}$$

36. ZTD:

37. **FHT liquidity measure:** Fong et al. (2017) propose a new bid-ask spread proxy based on the zeros measure in Lesmond et al. (1999). In their framework, symmetric transaction costs of $S/2$ leads to observed returns of

$$R = \begin{cases} R^* + \frac{S}{2} & \text{if } R^* < -\frac{S}{2}, \\ 0 & \text{if } -\frac{S}{2} \leq R^* < \frac{S}{2}, \\ R^* - \frac{S}{2} & \text{if } R^* \geq \frac{S}{2} \end{cases} \tag{18}$$

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

$$P_{FHT} = S \cdot 2 \cdot \sigma \cdot \Phi^{-1}\left(\frac{1 + P_{Zeros}}{2}\right), \tag{19}$$

38. Gibbs

39. **Amihud illiquidity LF:** Following Amihud (2002), the measure is motivated to cap-

ture the price impact and is defined as,

$$Amihud = \frac{1}{N}\sum_{d=1}^{N}(\frac{|r_{i,d}|}{Q_{i,d}}), \tag{20}$$

40. Amihud Roll

41. Amihud Gibbs

42. Amihud effective tick

43. Amihud FHT

44. Amihud HL spread

45. Amihud CHL

46. **Short-term reversal: ?** find that the past month's bond return has a significantly negative coefficient in predicting this month return.

47. **Bond variance:** the variance of bond returns using a 36-month rolling window for each bond in our sample.

48. **Skewness:** Similar to the construction of variance, skewness is estimated using a 36-month rolling window for each bond in our sample.

$$SKEW_{i,t} = \frac{1}{n}\sum_{t=1}^{n}(\frac{R_{i,t} - \bar{R}_i}{\sigma_{i,t}})^3, \tag{21}$$

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

$$R_{i,t} = \alpha_i + \beta_i R_{m,t} + \gamma_i R_{m,t}^2 + \epsilon_{i,t}, \tag{22}$$

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

50. **Idiosyncratic skewness:** The idiosyncratic skewness of bond i is defined as the skewness of the residuals in co-skewness regression equation.

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

$$KURT_{i,t} = \frac{1}{n}\sum_{t=1}^{n}(\frac{R_{i,t} - \bar{R}_i}{\sigma_{i,t}})^4 - 3, \tag{23}$$

52. **VIX beta:** Following Chung et al. (2019), we estimate the following bond-level regression

$$R_{i,t} = \alpha_i + \beta_{1,i}MKT_t + \beta_{2,i}SMB_t + \beta_{3,i}HML_t + \beta_{4,i}DEF_t + \beta_{5,i}TERM_t + \beta_{6,i}\delta VIX_t + \epsilon_{i,t}, \tag{24}$$

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

53. **Macroeconomic uncertainty beta:** Following Bali et al. (2021b), for each bond-month in our sample, we estimate the uncertainty beta from monthly rolling regressions of excess bond returns on the change in the economic uncertainty index (δ UNC) and the excess bond market returns (MKT), using the past 24 to 36 months of data (as available):

54. **Bond market beta:** We estimate the bond market beta for each bond from the time-series regressions of individual bond excess returns on the bond market excess returns (MKTBond) using a 36-month rolling window. We compute the bond market excess return (MKTBond) as the value-weighted average returns of all corporate bonds in our sample minus the one-month Treasury-bill rate.

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

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

57. **Bond 6-month momentum:** Jostova et al. (2013) find that winners in the corporate bond market over the past six months outperform losers by 37 basis points (bps) per month. Accordingly, we define 6-month bond momentum as the cumulative 6-month return on the corporate bond skipping the most recent month.

58. **Bond 11-month momentum:** We define 11-month bond momentum as the cumulative 11-month return on the corporate bond skipping the most recent month.

59. **Long-term reversal:** Bali et al. (2021b) show that contrarian strategies based on long-term returns are statistically and economically profitable in the corporate bond market. Following their definition, we construct long-term bond reversal as cumulative returns from month $t-48$ to $t-13$.

60. Idiosyncratic bond volatility: Chung et al. (2019) investigate the cross-sectional relation between expected bond returns and idiosyncratic volatility. We follow their approach by measuring idiosyncratic volatility by the standard deviation of return residuals from a factor model, which includes Fama and French (1993) five factors as well as VIX.

61. The Hasbrouck measure for bid-ask spreads: Hasbrouck (2009) develops a Gibbs sampler estimation of the extended Roll model,

$$r_t = c \cdot \Delta D_t + \beta r_t^M + \epsilon_t, \tag{25}$$

where D_t is a sell side indicator, c is half of the effective bid-ask spread, and r_t^M denotes the market return on day t. By making inference of the latent D_t with Gibbs sampling, this estimator overcomes the negative spread estimates. We estimate Eq. (26) on a monthly basis and define $Hasbrouck = 2\hat{c}$.

62. Bloomberg quoted bid-ask spreads: We use quoted bid-ask spreads from the Bloomberg Generic Quote (BGN). Let B_t and A_t be the bid and ask quote for a given bond and

day t .We get our daily relative spread estimates

$$BGN_t = \frac{2(A_t - B_t)}{B_t + A_t} \tag{26}$$

63. The LOT measure and the percentage of zero returns: Lesmond et al. (1999) develop an effective spread estimator (LOT) based on the idea that an observed return is different from zero only if the true return exceeds the trading costs; Chen et al. (2007) apply this estimator to measuring corporate bond market illiquidity. The key assumption underlying this estimator is that the observed bond return r_t has the following relation with the unobserved "true return" r_t^*,

$$r_t = \begin{cases} r_t^* - \alpha_1 & \text{if } r_t^* < \alpha_1 \\ 0 & \text{if } \alpha_1 < r_t^* < \alpha_2 \\ r_t^* - \alpha_2 & \text{if } r_t^* > \alpha_2 \end{cases}, \tag{27}$$

where α_1 denotes the percentage transaction cost of selling the bond and α_2 the percentage transaction cost of buying it. Following the original LOT procedures, we impose a one-factor structure on r_t^*

$$r_t^* - rf_t = \beta(r_t^M - rf_t) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma),$$

and estimate model parameters $\{\alpha_1, \alpha_2, \beta, \sigma\}$ with a maximum likelihood estimator.[7] It follows that the measure for effective bid-ask spread is given by

$$LOT = \alpha_2 - \alpha_1.$$

As discussed in Chen et al. (2007), a closely related measure is Percentage Zeros (ZRD), which denotes the proportion of days observing a zero return in the month.

64. Equity-Debt-Spread: Correia et al. (2012) define value signals as

$$CRV_{i,t} = \widetilde{CS}_{i,t} - CS_{i,t},$$

where $\widetilde{CS}_i$ and CS_i denote the modelled and observed yield spread of bond i, and the former is calculated with structural-model-implied default probability. Since $\widetilde{CS}_i$

[7] Regarding the definition of three different regions in the likelihood function, we follow Chen et al. (2007) by breaking out the three regions solely based on the observed bond returns. It corresponds to the "LOT Y-split" measure as defined in Goyenko et al. (2009).

is estimated with pricing information in the equity market, CRV is indeed closely related to the debt-equity-spread as studied by Chen et al. (2025). We adopt their terminology and label this covariate as equity-debt-spread.

65. Relative value: The relative value is defined as the difference between the quote yield and the fair value yield of a bond. The fair value yield is the fitted yield implied by issuer-specifc yield curves, which are constructed by the Nelson-Siegel model exploiting the price information of bonds issued by similar firms. Specifically, the yield curve of a firm is modelled as

$$y_{it}(\tau) = f^1(X_{it}) + \frac{1 - e^{-\lambda}\tau}{\lambda\tau} f^2(X_{it}) + (\frac{1 - e^{-\lambda}\tau}{\lambda\tau} - e^{-\lambda}\tau) f^3(X_{it})$$

where X_{it} is firm characteristics, τ is the maturity, and f^1, f^2, f^3 are the level, slope and curvature factors in the Nelson-Siegel model. The function forms of three factors are estimated to minimize pricing error via kernel smoothing.

A.2 Equity Characteristics

Table 2: Stock Characteristics

Accruals	
1. Total accruals	Richardson et al. (2005)
2. Change in current operating working capital	Richardson et al. (2005)
3. Operating accruals	Sloan (1996)
4. Percent operating accruals	Hafzalla et al. (2011)
5. Percent total accruals	Hafzalla et al. (2011)
6. Years 16-20 lagged returns, nonannual	Heston and Sadka (2008)
Debt issuance	
7. Abnormal corporate investment	Titman et al. (2004)
8. Growth in book debt (3 years)	Lyandres et al. (2008)
9. Change in financial liabilities	Richardson et al. (2005)
10. Change in noncurrenct operating liabilities	Richardson et al. (2005)
11. Change in net financial assets	Richardson et al. (2005)
12. Earnings persistence	Francis et al. (2004)
13. Net operating assets	Hirshleifer et al. (2004)
Investment	
14. Liquidity of book assets	Ortiz-Molina and Phillips (2014)
15. Asset growth	Cooper et al. (2008)
16. CAPEX growth (1 year)	Xie (2001)
17. CAPEX growth (2 years)	Anderson and Garcia-Feijoo (2006)
18. CAPEX growth (3 years)	Anderson and Garcia-Feijoo (2006)
19. Change in common equity	Richardson et al. (2005)
20. Change in current operating assets	Richardson et al. (2005)
21. Change in current operating liabilities	Richardson et al. (2005)
22. Change in noncurrent operating assets	Richardson et al. (2005)
23. Change in net noncurrent operating assets	Richardson et al. (2005)
24. Hiring rate	Belo et al. (2014)
25. Inventory growth	Belo and Lin (2012)
26. Inventory change	Thomas and Zhang (2002)
27. Change in long-term net operating assets	Fairfield et al. (2003)
28. Mispricing factor: Management	Stambaugh and Yuan (2017)
29. Change in net operating assets	Hirshleifer et al. (2004)
30. Change in PPE and inventory	Lyandres et al. (2008)
31. Long-term reversal	De Bondt and Thaler (1985)
32. Sales growth (1 year)	Lakonishok et al. (1994)
33. Sales growth (3 years)	Lakonishok et al. (1994)
34. Sales growth (1 quarter)	
35. Years 2-5 lagged returns, nonannual	Heston and Sadka (2008)

This table lists the firm characteristics. We adopt 153 stock characteristics from Jensen et al. (2023).

Low leverage	
36. Firm age	Jiang et al. (2005)
37. Liquidity of market assets	Ortiz-Molina and Phillips (2014)
38. Book leverage	Fama and French (1992)
39. The high-low bid-ask spread	Corwin and Schultz (2012)
40. Cash-to-assets	Palazzo (2012)
41. Net debt-to-price	Penman et al. (2007)
42. Earnings volatility	Francis et al. (2004)
43. R&D-to-sales	Chan et al. (2001)
44. R&D capital-to-book assets	Li (2011)
45. Asset tangibility	Hahn and Lee (2009)
46. Altman Z-score	Dichev (1998)
Low risk	
47. Market Beta	Fama and MacBeth (1973)
48. Dimson Beta	Dimson (1979)
49. Frazzini-Pedersen market beta	Frazzini and Pedersen (2014)
50. Downside beta	Ang et al. (2006)
51. Earnings variability	Francis et al. (2004)
52. Idiosyncratic volatility from the CAPM (21 days)	
53. Idiosyncratic volatility from the CAPM (252 days)	Ali et al. (2003)
54. Idiosyncratic volatility from the Fama-French 3-factor model	Ang et al. (2006)
55. Idiosyncratic volatility from the q-factor model	
56. Return volatility	Ang et al. (2006)
57. Cash flow volatility	Huang (2009)
58. Maximum daily return	Bali et al. (2011)
59. Highest 5 days of return	Bali et al. (2017)
60. Years 6-10 lagged returns, nonannual	Heston and Sadka (2008)
61. Share turnover	Datar et al. (1998)
62. Number of zero trades with turnover as tiebreaker (1 month)	Liu (2006)
63. Number of zero trades with turnover as tiebreaker (6 months)	Liu (2006)
64. Number of zero trades with turnover as tiebreaker (12 month2)	Liu (2006)
Momentum	
65. Current price to high price over last year	George and Hwang (2004)
66. Residual momentum t-6 to t-1	Blitz et al. (2011)
67. Residual momentum t-12 to t-1	Blitz et al. (2011)
68. Price momentum t-3 to t-1	Jegadeesh and Titman (1993)
69. Price momentum t-6 to t-1	Jegadeesh and Titman (1993)
70. Price momentum t-9 to t-1	Jegadeesh and Titman (1993)
71. Price momentum t-12 to t-1	Jegadeesh and Titman (1993)
72. Year 1-lagged returns, nonannual	Heston and Sadka (2008)
Profit growth	
73. Change sales minus change Inventory	Abarbanell and Bushee (1998)
74. Change sales minus change receivables	Abarbanell and Bushee (1998)
75. Change sales minus change SG&A	Abarbanell and Bushee (1998)
76. Change in quarterly return on assets	
77. Change in quarterly return on equity	
78. Standardized earnings surprise	Foster et al. (1984)
79. Change in operating cash flow to assets	Bouchaud et al. (2019)
80. Price momentum t-12 to t-7	Novy-Marx (2012)
81. Labor force efficiency	Abarbanell and Bushee (1998)

82. Standardized revenue surprise	Jegadeesh and Livnat (2006)
83. Year 1-lagged return, annual	Heston and Sadka (2008)
84. Tax expense surprise	Thomas and Zhang (2011)
Profitability	
85. Coefficient of variation for dollar trading volume	Chordia et al. (2001)
86. Coefficient of variation for share turnover	Chordia et al. (2001)
87. Return on net operating assets	Soliman (2008)
88. Profit margin	Soliman (2008)
89. Pitroski F-score	Pitroski (2000)
90. Return on equity	Haugen and Baker (1996)
91. Quarterly return on equity	Hou et al. (2015)
92. Ohlson O-score	Dichev (1998)
93. Operating cash flow to assets	Bouchaud et al. (2019)
94. Operating profits-to-book equity	Fama and French (2015)
95. Operating profits-to-lagged book equity	
Quality	
96. Capital turnover	Haugen and Baker (1996)
97. Cash-based operating profits-to-book assets	
98. Cash-based operating profits-to-lagged book assets	Ball et al. (2016)
99. Operating profits-to-book assets	
100. Operating profits-to-lagged book assets	Ball et al. (2016)
101. Change gross margin minus change sales	Abarbanell and Bushee (1998)
102. Gross profits-to-assets	Novy-Marx (2013)
103. Gross profits-to-lagged assets	
104. Mispricing factor: Performance	Stambaugh and Yuan (2017)
105. Number of consecutive quarters with earnings increases	Barth et al. (1999)
106. Quarterly return on assets	Balakrishnan et al. (2010)
107. Operating leverage	Novy-Marx (2011)
108. Quality minus Junk: Composite	Asness et al. (2019)
109. Quality minus Junk: Growth	Asness et al. (2019)
110. Quality minus Junk: Profitability	Asness et al. (2019)
111. Quality minus Junk: Safety	Asness et al. (2019)
112. Assets turnover	Soliman (2008)
Seasonality	
113. Market correlation	Asness et al. (2020)
114. Coskewness	Harvey and Siddique (2000)
115. Net debt issuance	Bradshaw et al. (2006)
116. Kaplan-Zingales index	Lamont et al. (2001)
117. Change in long-term investments	Richardson et al. (2005)
118. Change in short-term investments	Richardson et al. (2005)
119. Taxable income-to-book income	Lev and Nissim (2004)
120. Years 2-5 lagged returns, annual	Heston and Sadka (2008)
121. Years 6-10 lagged returns, annual	Heston and Sadka (2008)
122. Years 11-15 lagged returns, annual	Heston and Sadka (2008)
123. Years 11-15 lagged returns, nonannual	Heston and Sadka (2008)
124. Years 16-20 lagged returns, annual	Heston and Sadka (2008)

Size	
125. Amihud measure	Amihud (2002)
126. Dollar trading volume	Brennan et al. (1998)
127. Market equity	Banz (1981)
128. Price per share	Miller and Scholes (1982)
129. R&D-to-market	Chan et al. (2001)
Short-term reversal	
130. Idiosyncratic skewness from the CAPM	
131. Idiosyncratic skewness from the Fama-French 3-factor model	Bali et al. (2016)
132. Idiosyncratic skewness from the q-factor model	
133. Total skewness	Bali et al. (2016)
134. Short-term reversal	Jegadeesh (1990)
135. Highest 5 days of return scaled by volatility	Asness et al. (2020)
Value	
136. Assets-to-market	Fama and French (1992)
137. Book-to-market equity	Rosenberg et al. (1985)
138. Book-to-market enterprise value	Penman et al. (2007)
139. Net stock issues	Pontiff and Woodgate (2008)
140. Debt-to-market	Bhandari (1988)
141. Dividend yield	Litzenberger and Ramaswamy (1979)
142. Ebitda-to-market enterprise value	Loughran and Wellman (2011)
143. Equity duration	Dechow et al. (1999)
144. Net equity issuance	Bradshaw et al. (2006)
145. Net total issuance	Bradshaw et al. (2006)
146. Equity net payout	Daniel and Titman (2006)
147. Net payout yield	Boudoukh et al. (2007)
148. Payout yield	Boudoukh et al. (2007)
149. Free cash flow-to-price	Asness et al. (2019)
150. Intrinsic value-to-market	Frankel and Lee (1998)
151. Earnings-to-price	Basu (1983)
152. Operating cash flow-to-market	Desai et al. (2004)
153. Sales-to-market	Barbee Jr et al. (1996)

Appendix B Machine Learning Methods

The excess return of a corporate bond i at time $t+1$ can be described as:

$$r_{i,t+1} = E_t(r_{i,t+1}) + \epsilon_{i,t+1} \tag{28}$$

where

$$E_t(r_{i,t+1}) = g^*(z_{i,t}) \tag{29}$$

is the time t expected return that is represented by a flexible function g^* of corporate bond i's P-dimensional characteristics $z_{it} = (z_{i,1,t}, z_{i,2,t}, \cdots, z_{i,P,t})'$. For ease of presentation, we assume a balanced panel of corporate bonds. We index corporate bonds by $i = 1, \cdots, N_t$ and months by $t = 1, \cdots, T$, where N is the number of corporate bonds at time t.

B.1 Linear Regression

The least complex but straightforward method to approximate g^* is a simple linear function

$$g^*(z_{i,t}; \theta) = z'_{i,t}\theta \tag{30}$$

where $\theta = (\theta_1, \cdots, \theta_P)$ are P-dimensional model parameters that can be estimated via the ordinary least square (OLS). The objective function is

$$\min_{\theta} \mathcal{L}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(r_{i,t+1} - g(z_{i,t}; \theta)\right)^2 \tag{31}$$

Apart from considering the raw predictor variables as $z_{i,t}$, we also cross-sectionally rank all characteristics period-by-period and map these ranks into the interval $[-1, 1]$ to mitigate the effect of outliers.

B.2 Penalized Linear: LASSO, Ridge, and Elastic Net

When the number of predictors P approaches the number of observations T, the OLS estimator often becomes inefficient or even inconsistent. A popular strategy to avoid overfitting and improve the model interpretability is to append a penalty term to the objective function

(31). In its general form, the objective function of a penalized regression can be written as

$$\mathcal{L}(\theta; \cdot) = \mathcal{L}(\theta) + \phi(\theta; \cdot) \tag{32}$$

where $\phi(\theta; \cdot)$ is the penalty function of θ.The constraints on the coefficients allow that the coefficients of those less relevant variables are shrunk towards zero or exactly zero, which improves the model's out-of-sample stability.

A widely-used penalty function in the machine learning literature is the so-called "elastic net" (Zou and Hastie, 2005), which takes the form

$$\phi(\theta; \lambda, \rho) = \lambda(1-\rho)\sum_{j=1}^{P}|\theta_j| + \frac{1}{2}\lambda\rho\sum_{j=1}^{P}\theta_j^2 \tag{33}$$

The elastic net involves two non-negative hyperparameters, λ and ρ. The $\rho = 0$ case corresponds to LASSO where the penalty term is an l_1 regularization and it sets some coefficients exactly to zero. In this sense, it is a variable selection method that imposes model sparsity. The $\rho = 1$ case corresponds to the ridge regression where the penalty term is an l_2 regularization. In contrast to LASSO, the ridge regression shrinks all coefficient estimates closer to zero but never exactly imposes zeros. It is a shrinkage method that helps prevent coefficients from becoming unduly large in magnitude. The elastic net is a combination of the two by setting ρ between 0 and 1.

B.3 Dimension Reduction: PCR and PLS

Penalized linear models can produce suboptimal forecasts when predictors are highly correlated. Dimension reduction techniques such as principal components regression (PCR) and partial least squares help de-correlate highly dependent predictors and better isolate the signal in predictors. The linear regression model can be written in matrix form as

$$R = Z\theta + E \tag{34}$$

where R is the $NT \times 1$ vector of $r_{i,t+1}$, Z is the $NT \times P$ matrix of stacked predictors $z_{i,t}$, and E is a $NT \times 1$ vector of residual $\epsilon_{i,t+1}$. PCR and PLS aim to project the set of predictors from dimension P to a much smaller number of K-dimensional space. The forecasting model thus becomes

$$R = (Z\Omega_K)\theta_K + \tilde{E} \tag{35}$$

where Ω_K is $P \times K$ matrix with columns $w_1, w_2, \cdots, w_K$. Each w_j is the set of linear combination weights used to create the jth predictive component.

PCR and PLS differ from their choices for the combination weights. PCR maximizes the common variation across the predictors and the jth principal component solves

$$w_j = \arg\max Var(Zw) \quad \text{s.t.} \quad w'w = 1, \quad Cov(Zw, Zw_l) = 0, \quad l = 1, 2, \cdots, j-1 \tag{36}$$

The target variable is disgarded when extracting the principal components thus there is no guarantee that they are the best latent factors for return prediction. In contrast, PLS links the target variable to return predictors and seeks K linear combinations of Z that have maximal predictive association with asset returns. The weights used to construct the jth PLS component solve

$$w_j = \arg\max Cov^2(R, Zw) \quad \text{s.t.} \quad w'w = 1, \quad Cov(Zw, Zw_l) = 0, \quad l = 1, 2, \cdots, j-1 \tag{37}$$

For both methods, K is a hyperparameter can be tuned from the validation sample.

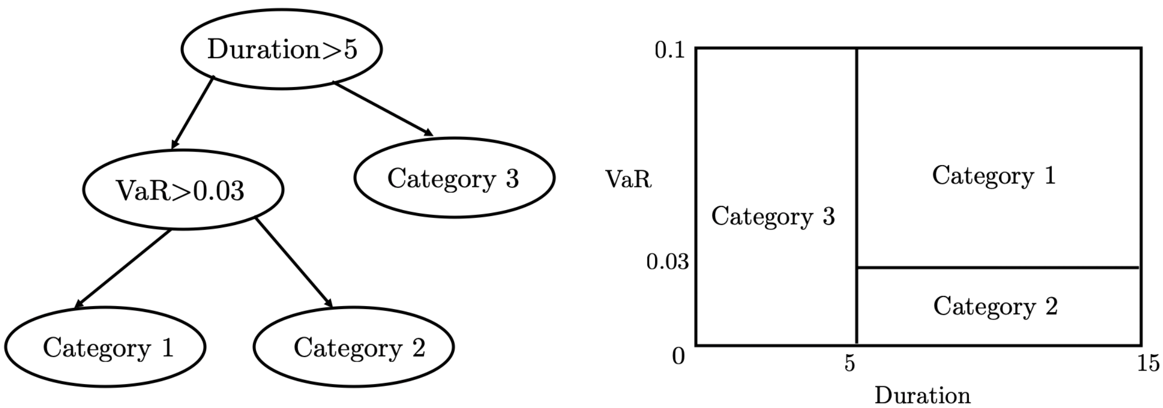
Penalized linear regression and dimension reduction techniques do not account for nonlinear relations. To address this issue, we consider two classes of popular nonlinear methods: regression trees and neural networks.

B.4 Regression Trees: XGBoost, Random Forests, and Extra-Trees

Regression trees are a kind of simple yet powerful non-parametric machine learning methods. A tree “grows” in a sequence of steps. At each step, a new “branch” splits the data leftover from the preceding step into bins based on one of the predictors. This sequential branching slices the input space into rectangular partition, and approximate g^* with a simple model (such as the average value of the outcome variable) in each partion.

Figure A1 exhibits an example of a regression tree with two predictors, “Duration”, and “VaR”. The left panel describes how the tree assigns each observation in the sample to a partition based on its predictor values. First, observations are splitted on Duration. Observations with Duration above the breakpoint of 5 are assigned to Category 3. Corporate bonds with shorter duration are further splitted according to their VaR. Observations with shorter duration and VaR larger than 0.03 are partitioned into Category 1, while the rest go into Category 2.

Figure A1: Example of a regression tree



Mathematically, the prediction of a tree, $\mathcal{T}$, with K "leaves" (terminal nodes), and depth L can be formed as

$$g(z_{i,t}; \theta, K, L) = \sum_{k=1}^{K} \theta_k \mathbf{1}_{\{z_{i,t} \in C_k(K)\}} \tag{38}$$

where $C_k(L)$ is one of the K partion of the data, $\mathbf{1}_{\{\cdot\}}$ is an indicator function, and θ_k is simply the mean of outcome variables within the partition C_k.

A very large tree is prone to overfitting, while a small tree might not capture the important predictors thus have weak prediction accuracy. Here, we consider three "ensemble" methods that combine forecasts from many trees into a single one to improve the prediction accuracy and mitigate overfitting.

The first one is boosting. After fitting a shallow tree with depth L, a second simple tree (with the same depth) is fitted to the residuals from the first tree. Forecasts from these two trees are then added together to form an ensemble prediction of the outcome and update the prediction residuals. The forecast component from the second tree is shrunken by a factor $\nu \in (0, 1)$ to prevent overfitting and allow more and different shaped trees. This procedure is iterated until there are a total number of B trees in the ensemble. The final output can be represented by an additive model of shallow trees with tuning parameters (L, ν, B). Specifically, we adpot XGboost in this paper, which is a powerful and efficient machine learning algorithm based on the gradient boosting framework.

The second one is random forests, a refinement of a more general procedure known as or bootstrap aggregation, or "bagging" (Breiman, 2001). The baseline tree bagging procedure is to draw B bootstrapped samples of the data and build a separate regression tree using each sample. The resulting predictions are then averaged to produce a single forecast. Unlike boosting, each individual tree is deep and not pruned thus makes the individual prediction variable. Averaging over multiple predictions can reduce the variance and improve the prediction accuracy impressively.

Random forests provide an improvement over bagged trees by way of a small tweak that decorrelates the trees. As in bagging, a number of regression trees are built on bootstrapped training samples. But when building these trees, only a random sample of m predictors is chosen as split candidates each time a split in a tree is considered. By considering only a subset of the predictors, random forecasts overcome the problem that the predictions from the bagged trees will be highly correlated and there is no large reduction in variance if there are strong predictors in the sample.

The third one is extra trees, which is also a model based on bagging. Unlike traditional decision tree algorithms, extra trees introduces additional randomness by selecting split thresholds randomly for each feature, rather than searching for the most optimal split. This approach leads to greater tree diversity and often improves generalization performance. By averaging predictions from many such randomized trees, extra trees can reduce overfitting and increase robustness, making it a powerful and efficient alternative to random forests.

Table 1: Summary Statistics on Bond Returns and Characteristics

	N_obs	Mean	Std	Percentiles				
				5th	25th	50th	75th	95th
Excess return (%)	1,261,971	0.40	2.60	−3.30	−0.50	0.30	1.30	4.10
Price return (%)	1,261,971	0.10	2.50	−3.50	−0.70	0.10	1.10	3.80
Yield implied carry (%)	1,261,971	0.40	0.20	0.10	0.30	0.40	0.50	0.70
Duration	1,261,971	6.26	3.92	1.33	3.16	5.39	8.68	13.80
Rating	1,261,724	7.21	3.11	2	5	7	9	13
Age	1,261,075	4.23	3.95	0.29	1.44	3.13	5.82	12.41
Maturity	1,261,971	9.88	8.85	1.39	3.55	6.70	13.26	28.10
Size	1,261,971	11.85	2.01	7.71	11.08	12.43	13.12	14.22

The table reports the descriptive statistics for excess bond returns, price returns, yield implied carry, and some characteristics (including credit rating, duration, size, maturity, and age). The yield-implied carry is defined in Eq. (2). The sample period spans from January 1991 to June 2019.

Table 2: Out-of-Sample Security-Level Prediction Performance

	Obs	OLS	Lasso	Ridge	Enet	PCR	PLS	XGBoost	RF	ExtraTrees	Comb	Yield-implied
Panel A: Bond and Equity Characteristics, Public Firms												
Total Return	721,785	2.14	2.02	2.10	2.24	1.16	2.08	1.69	1.98	1.87	2.29	2.95
Carry+Price Return	721,785	3.36	3.55	3.39	3.68	2.73	3.43	3.06	3.24	3.35	3.63	2.95
Panel B: Bond Characteristics, Private Firms												
Total Return	226,316	2.29	1.74	2.20	1.78	1.74	2.06	1.40	1.88	1.88	2.05	3.61
Carry+Price Return	226,316	3.54	3.41	3.49	3.42	3.28	3.29	3.21	3.30	3.33	3.48	3.61

The table summarizes the predictive ability of machine learning models for one-month-ahead corporate bond returns. It reports the out-of-sample R-squared

$$R^2_{OOS} = 1 - \frac{\sum (rx_{i,t} - \hat{rx}_{i,t})^2}{\sum rx^2_{i,t}},$$

as produced by ordinary least squares (OLS), Lasso, ridge regression (Ridge), elastic net (Enet), principal component regression (PCR), partial least squares (PLS), XGBoost, random forest (RF), Extratrees, as well as the forecast combining method (Comb). The last column reports the R^2_{OOS} generated by yield-implied expected bond returns, which is defined in Eq. (2). Results in Panel A are based on bond and stock characteristics among public firms, and Panel B only includes bond characteristics as bond return predictors among private firms. The sample period spans from July 2004 to June 2019.

Table 3: Out-of-Sample Security-Level Prediction Performance in Subsamples

	Obs	OLS	Lasso	Ridge	Enet	PCR	PLS	XGBoost	RF	ExtraTrees	Comb	Yield-implied
Panel A: Investment-Grade Bonds												
Bond and Equity Characteristics, Public Firms												
Total Return	574,787	1.85	2.04	1.84	2.07	1.03	1.94	1.58	1.70	1.70	2.19	2.04
Carry+Price Return	574,787	2.40	2.59	2.42	2.62	1.72	2.50	1.94	2.36	2.39	2.72	2.04
Bond Characteristics, Private Firms												
Total Return	184,359	1.85	1.61	1.79	1.58	1.46	1.66	1.24	1.56	1.61	1.78	2.88
Carry+Price Return	184,359	2.43	2.45	2.40	2.39	2.27	2.23	2.23	2.27	2.32	2.47	2.88
Panel B: High-Yield Bonds												
Bond and Equity Characteristics, Public Firms												
Total Return	146,998	2.54	2.00	2.45	2.48	1.32	2.26	1.85	2.36	2.11	2.43	4.19
Carry+Price Return	146,998	4.66	4.87	4.71	5.13	4.11	4.70	4.60	4.45	4.67	4.88	4.19
Bond Characteristics, Private Firms												
Total Return	41,802	3.01	1.95	2.88	2.11	2.20	2.73	1.67	2.41	2.31	2.49	4.81
Carry+Price Return	41,802	5.40	5.00	5.30	5.13	4.96	5.03	4.83	5.00	5.02	5.16	4.81

The table summarizes the predictive ability of machine learning models for one-month-ahead corporate bond returns among investment-grade and high-yield bonds. It reports the out-of-sample R-squared

$$R^2_{OOS} = 1 - \frac{\sum(rx_{i,t} - \hat{rx}_{i,t})^2}{\sum rx^2_{i,t}},$$

as produced by ordinary least squares (OLS), Lasso, ridge regression (Ridge), elastic net (ENet), principal component regression (PCR), partial least squares (PLS), XGBoost, random forest (RF), Extratrees, as well as the forecast combining method (Comb). The last column reports the R^2_{OOS} generated by yield-implied expected bond returns, which is defined in Eq. (2). Results in Panel A and Panel B are among investment-grade and high-yield bonds, respectively. The sample period spans from July 2004 to June 2019.

Table 4: Comparison of Out-of-Sample Prediction using Diebold-Mariano Tests

	OLS	Lasso	Ridge	Enet	PCR	PLS	XGBoost	RF	ExtraTrees	Comb
Panel A: Bond and Equity Characteristics, Public Firms										
ML Carry+Price Return VS ML Total Return	1.54	1.72*	1.60	1.64	1.60	1.75*	1.55	1.71*	1.97**	1.63
cryytm VS ML Total Return	0.93	1.07	0.97	0.84	1.00	1.83*	1.70*	0.89	1.13	0.78
Panel B: Bond Characteristics, Private Firms										
ML Carry+Price Return VS ML Total Return	1.72*	1.96*	1.75*	1.95*	1.69*	1.90*	2.14**	1.79*	1.75*	1.81*
Yield-implied VS ML Total Return	1.78*	2.20**	1.88*	2.14**	2.06**	2.21**	2.81***	1.96*	1.97*	1.94*

This table reports the paired Diebold-Mariano test statistics comparing the out-of-sample forecasting performance of convertible bonds between two models. In the first row, positive numbers indicate that the machine price return-target model outperforms the machine learning excess return target model. In the second row, positive numbers indicate that the yield-implied expected return model outperforms the machine learning excess return target model. Results in Panel A are based on bond and stock characteristics among public firms, and Panel B only includes bond characteristics as bond return predictors among private firms. *, **, and *** represent the 10%, 5%, and 1% significance levels, respectively. The sample period spans from July 2004 to June 2019.

Table 5: Performance of Hedged Portfolios Based on Return Forecasts

	Total Return					Carry+Price Return				
Model	Gross	Turnover	Adjusted for Hasbrouck	Alpha	Adjusted for Hasbrouck	Gross	Turnover	Adjusted for Hasbrouck	Alpha	Adjusted for Hasbrouck
Panel A: Bond and Stock Characteristics, Public Firms										
OLS	0.83	1.28	0.21	0.65	0.04	1.24	1.27	0.62	0.91	0.29
	(3.34)		(0.81)	(3.19)	(0.16)	(3.78)		(1.95)	(4.41)	(1.47)
Lasso	1.11	1.43	0.40	1.00	0.30	1.74	1.31	1.10	1.34	0.70
	(7.05)		(2.93)	(6.49)	(2.23)	(4.01)		(2.61)	(4.95)	(2.75)
Ridge	0.82	1.29	0.19	0.65	0.02	1.27	1.28	0.64	0.93	0.30
	(3.34)		(0.73)	(3.16)	(0.09)	(3.68)		(1.90)	(4.33)	(1.47)
Enet	1.09	1.39	0.41	0.97	0.29	1.63	1.32	0.99	1.27	0.63
	(6.80)		(2.82)	(6.37)	(2.18)	(3.95)		(2.48)	(4.89)	(2.59)
PCR	0.26	0.77	-0.10	0.19	-0.17	1.26	0.72	0.88	0.79	0.41
	(1.57)		(-0.60)	(1.36)	(-1.15)	(2.65)		(1.90)	(2.76)	(1.52)
PLS	0.85	1.22	0.25	0.66	0.07	1.32	1.23	0.70	0.99	0.38
	(5.06)		(1.46)	(5.13)	(0.48)	(3.73)		(2.09)	(4.36)	(1.84)
XGBoost	0.90	1.39	0.22	0.80	0.11	1.48	1.10	0.91	1.04	0.46
	(4.60)		(1.36)	(4.34)	(0.82)	(3.63)		(2.29)	(4.36)	(2.10)
RF	0.60	1.22	0.00	0.43	-0.17	1.18	1.03	0.64	0.78	0.24
	(2.93)		(0.01)	(2.50)	(-0.81)	(2.78)		(1.47)	(3.39)	(0.98)
ExtraTrees	0.62	1.21	0.05	0.45	-0.11	1.42	1.03	0.90	0.99	0.47
	(3.46)		(0.28)	(3.14)	(-0.66)	(3.43)		(2.21)	(4.13)	(2.06)
Comb	0.97	1.28	0.34	0.80	0.17	1.57	1.18	0.97	1.14	0.55
	(4.53)		(1.52)	(4.53)	(0.89)	(3.60)		(2.29)	(4.48)	(2.28)
Yield-implied	1.38	0.42	1.15	0.85	0.67	1.38	0.42	1.15	0.85	0.67
	(2.96)		(2.61)	(3.16)	(2.71)	(2.96)		(2.61)	(3.16)	(2.71)

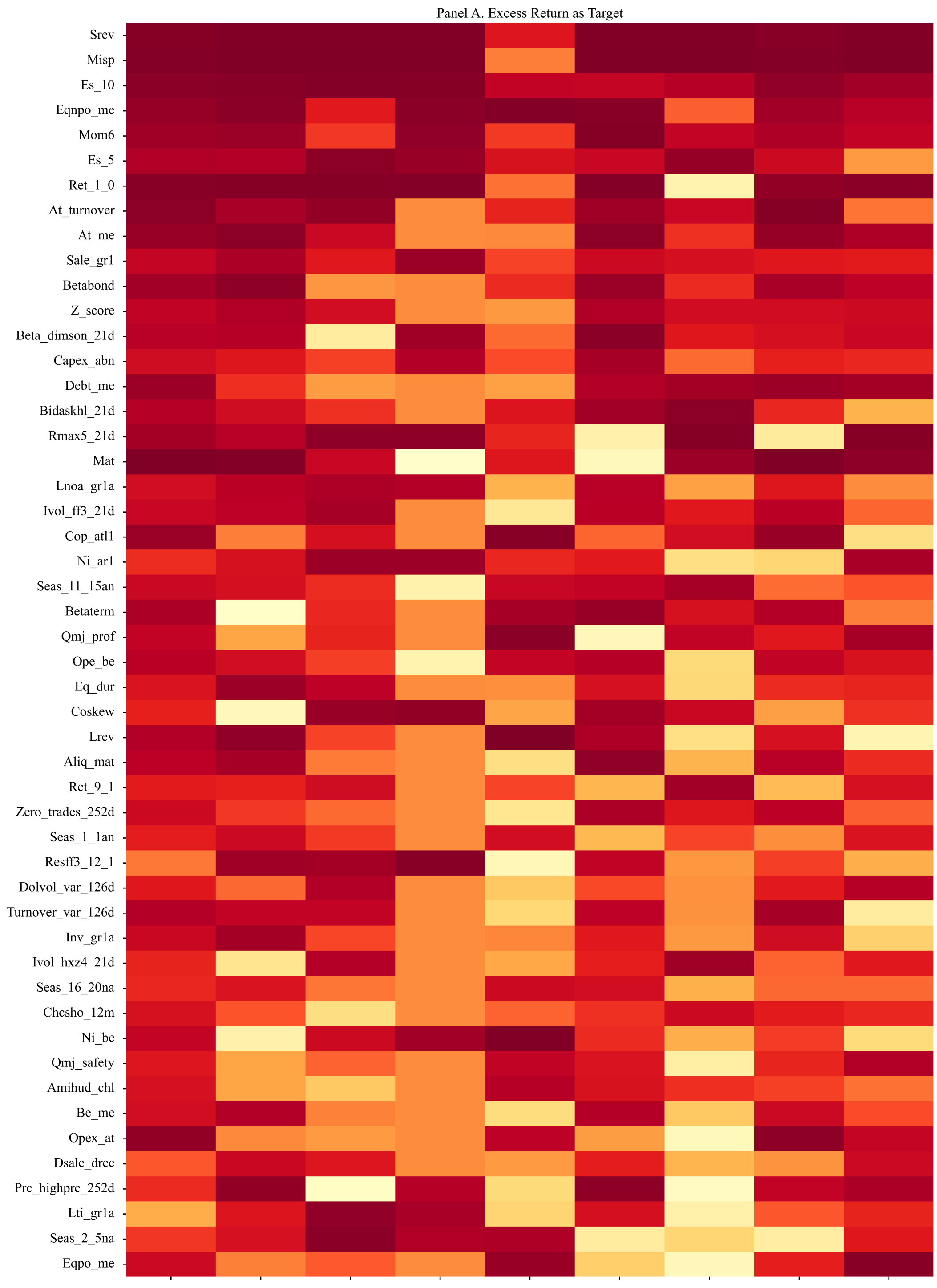
Table 5 (Continued)

Model	Total Return					Carry+Price Return				
	Gross	Turnover	Adjusted for Hasbrouck	Alpha	Adjusted for Hasbrouck	Gross	Turnover	Adjusted for Hasbrouck	Alpha	Adjusted for Hasbrouck
Panel B: Bond Characteristics, Private Firms										
OLS	0.81	1.09	0.30	0.54	0.02	1.02	1.10	0.49	0.67	0.13
	(3.38)		(1.29)	(2.93)	(0.14)	(3.44)		(1.66)	(3.37)	(0.70)
Lasso	0.78	1.15	0.25	0.58	0.06	1.32	0.99	0.82	0.88	0.37
	(3.30)		(1.18)	(2.58)	(0.31)	(3.89)		(2.39)	(4.19)	(1.81)
Ridge	0.83	1.11	0.32	0.56	0.03	1.03	1.12	0.50	0.68	0.14
	(3.43)		(1.34)	(2.99)	(0.20)	(3.45)		(1.67)	(3.41)	(0.72)
Enet	0.84	1.13	0.32	0.65	0.13	1.26	0.97	0.76	0.83	0.32
	(3.58)		(1.49)	(2.94)	(0.72)	(3.76)		(2.22)	(3.98)	(1.55)
PCR	0.64	0.88	0.27	0.26	-0.11	1.12	0.91	0.67	0.68	0.23
	(2.64)		(1.13)	(1.75)	(-0.71)	(3.26)		(1.93)	(3.37)	(1.11)
PLS	0.70	0.97	0.25	0.41	-0.05	0.94	0.88	0.53	0.56	0.14
	(2.92)		(1.08)	(2.32)	(-0.28)	(3.02)		(1.71)	(2.80)	(0.75)
XGBoost	0.30	1.17	-0.03	0.40	0.08	1.16	0.84	0.77	0.71	0.32
	(1.46)		(-0.17)	(2.06)	(0.38)	(3.34)		(2.22)	(3.39)	(1.53)
RF	0.47	1.16	-0.05	0.24	-0.27	0.98	0.87	0.56	0.59	0.16
	(1.97)		(-0.19)	(1.61)	(-1.87)	(2.95)		(1.66)	(3.03)	(0.85)
ExtraTrees	0.70	1.15	0.20	0.46	-0.04	1.04	0.93	0.60	0.63	0.18
	(3.13)		(0.91)	(2.54)	(-0.23)	(3.16)		(1.77)	(3.19)	(0.92)
Comb	0.81	1.05	0.33	0.52	0.03	1.14	0.96	0.67	0.70	0.23
	(3.21)		(1.32)	(2.92)	(0.20)	(3.35)		(1.96)	(3.40)	(1.12)
Yield-implied	1.22	0.42	1.15	0.76	0.67	1.22	0.42	1.15	0.76	0.67
	(3.42)		(2.61)	(3.43)	(2.71)	(3.42)		(2.61)	(3.43)	(2.71)

This table describes the profitability of forecast-based long-short portfolios, which are longing in the 10th decile with the highest predicted return and shorting in the 1st decile with the lowest predicted return. It reports the results on the average monthly excess returns and risk-adjusted returns based on the market factor. Columns labelled "Gross" present returns/alphas unadjusted for trading costs, and columns labelled "Hasbrouck" report returns/alphas adjusted for trading costs. Trading costs are calculated as the product of the portfolio turnover and the average security-level transaction cost, with the latter measured by the Hasbrouck measure. In parentheses are t-statistics computed with Newey-West error correction (using 5 lags). Results in Panel A are based on bond and stock characteristics among public firms, and Panel B only includes bond characteristics as bond return predictors among private firms. The sample period spans from July 2004 to December 2019.

Figure 1: Variable Importance for Public Firms

This figure shows the ranking of the bond and stock characteristics regarding overall model contribution using the R^2 measure. Features are sorted according to the sum of their rankings across all models, with the most influential features at the top and the least influential at the bottom. The columns correspond to the individual models, and the color gradient in each column indicates the variables from most influential (dark orange) to least influential (white). The top 50 features are shown in the figure due to space constraints.



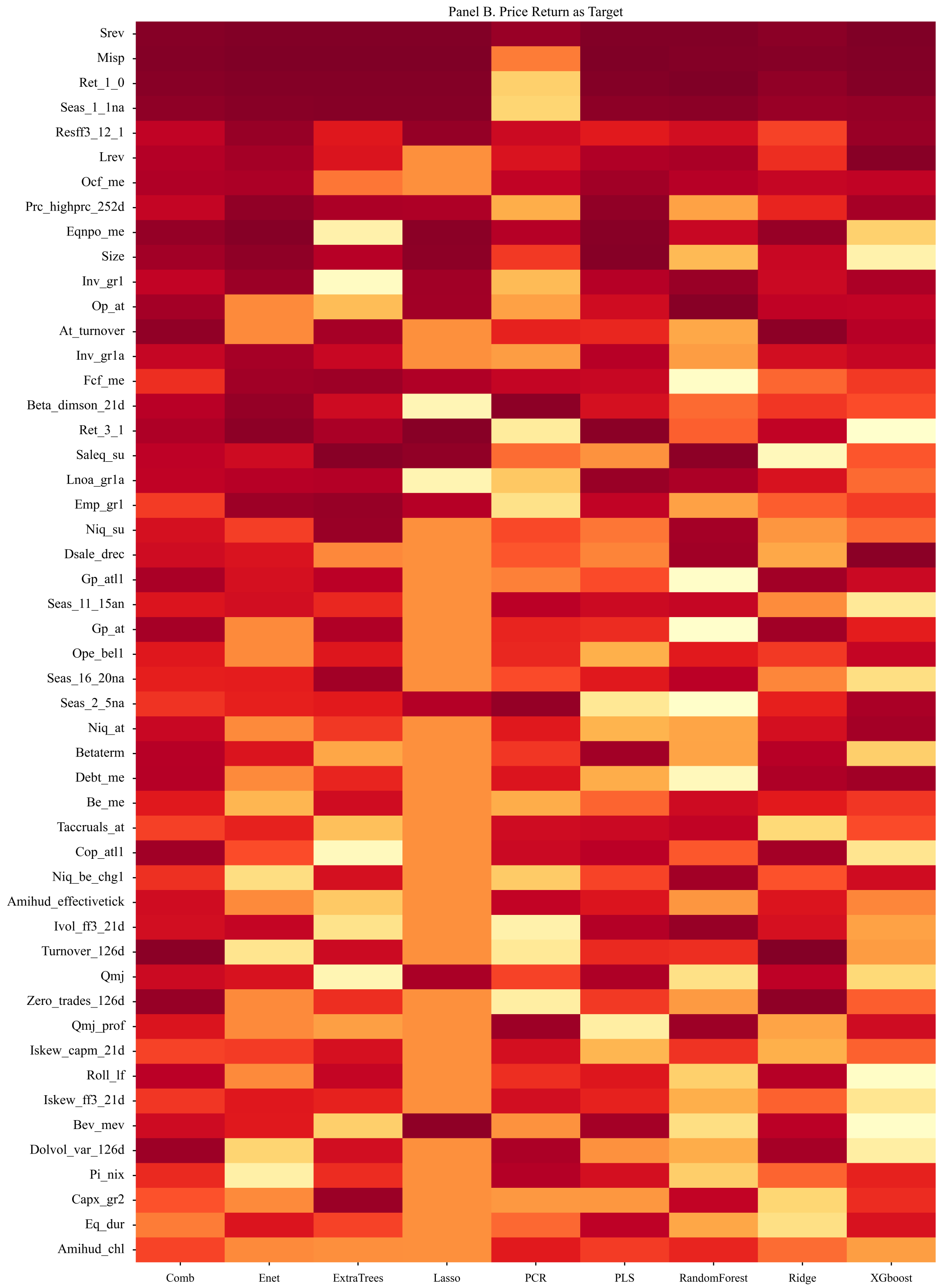
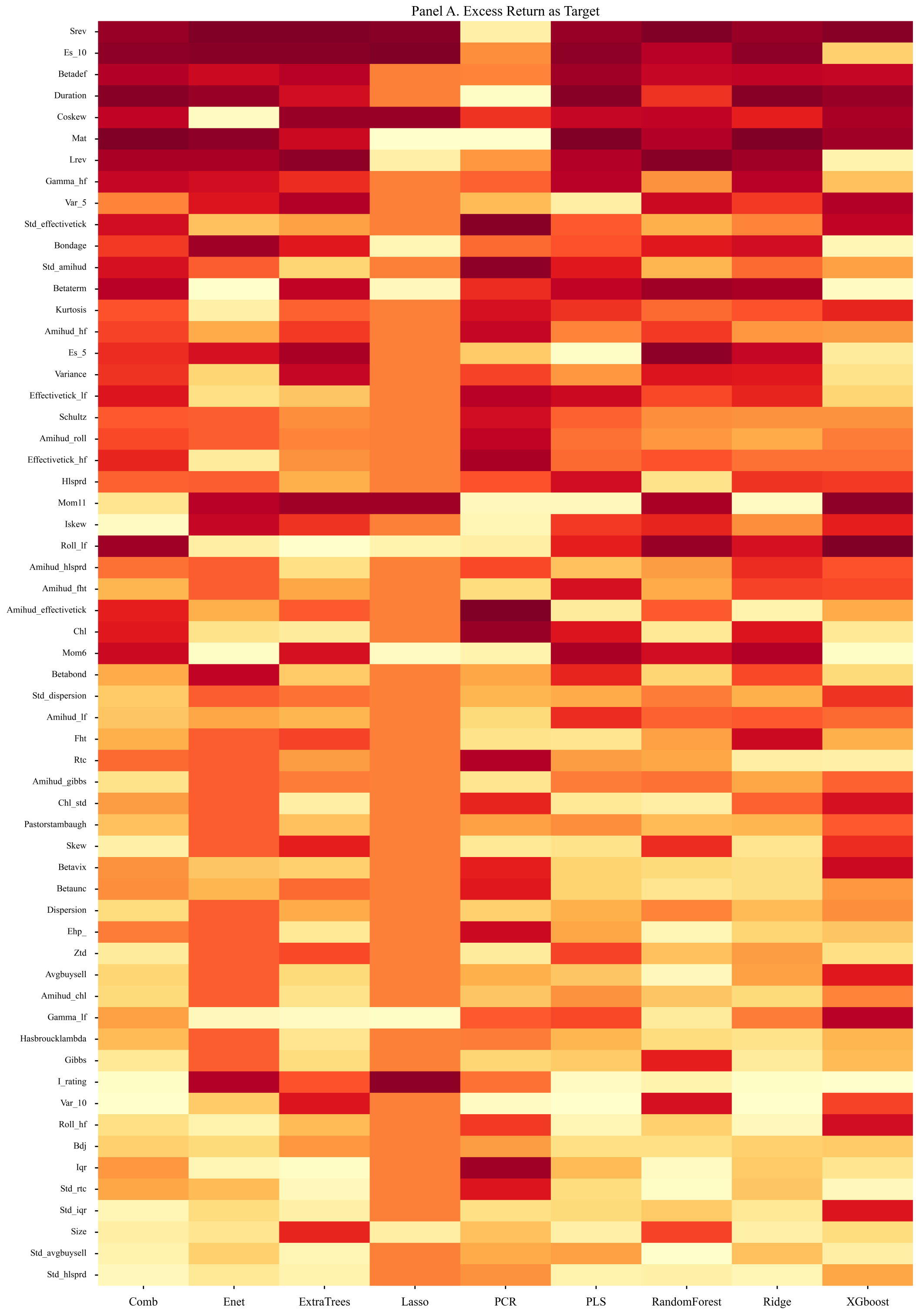

Panel B. Price Return as Target
Srev
Misp
Ret_1_0
Seas_1_1na
Resff3_12_1
Lrev
Ocf_me
Prc_highprc_252d
Eqnpo_me
Size
Inv_gr1
Op_at
At_turnover
Inv_gr1a
Fcf_me
Beta_dimson_21d
Ret_3_1
Saleq_su
Lnoa_gr1a
Emp_gr1
Niq_su
Dsale_drec
Gp_atl1
Seas_11_15an
Gp_at
Ope_bel1
Seas_16_20na
Seas_2_5na
Niq_at
Betaterm
Debt_me
Be_me
Taccruals_at
Cop_atl1
Niq_be_chg1
Amihud_effectivetick
Ivol_ff3_21d
Turnover_126d
Qmj
Zero_trades_126d
Qmj_prof
Iskew_capm_21d
Roll_lf
Iskew_ff3_21d
Bev_mev
Dolvol_var_126d
Pi_nix
Capx_gr2
Eq_dur
Amihud_chl
Comb
Enet
ExtraTrees
Lasso
PCR
PLS
RandomForest
Ridge
XGboost

Figure 2: Variable Importance for Private Firms

This figure shows the ranking of the bond and stock characteristics regarding overall model contribution using the R^2 measure. Features are sorted according to the sum of their rankings across all models, with the most influential features at the top and the least influential at the bottom. The columns correspond to the individual models, and the color gradient in each column indicates the variables from most influential (dark orange) to least influential (white).



Panel B. Price Return as Target

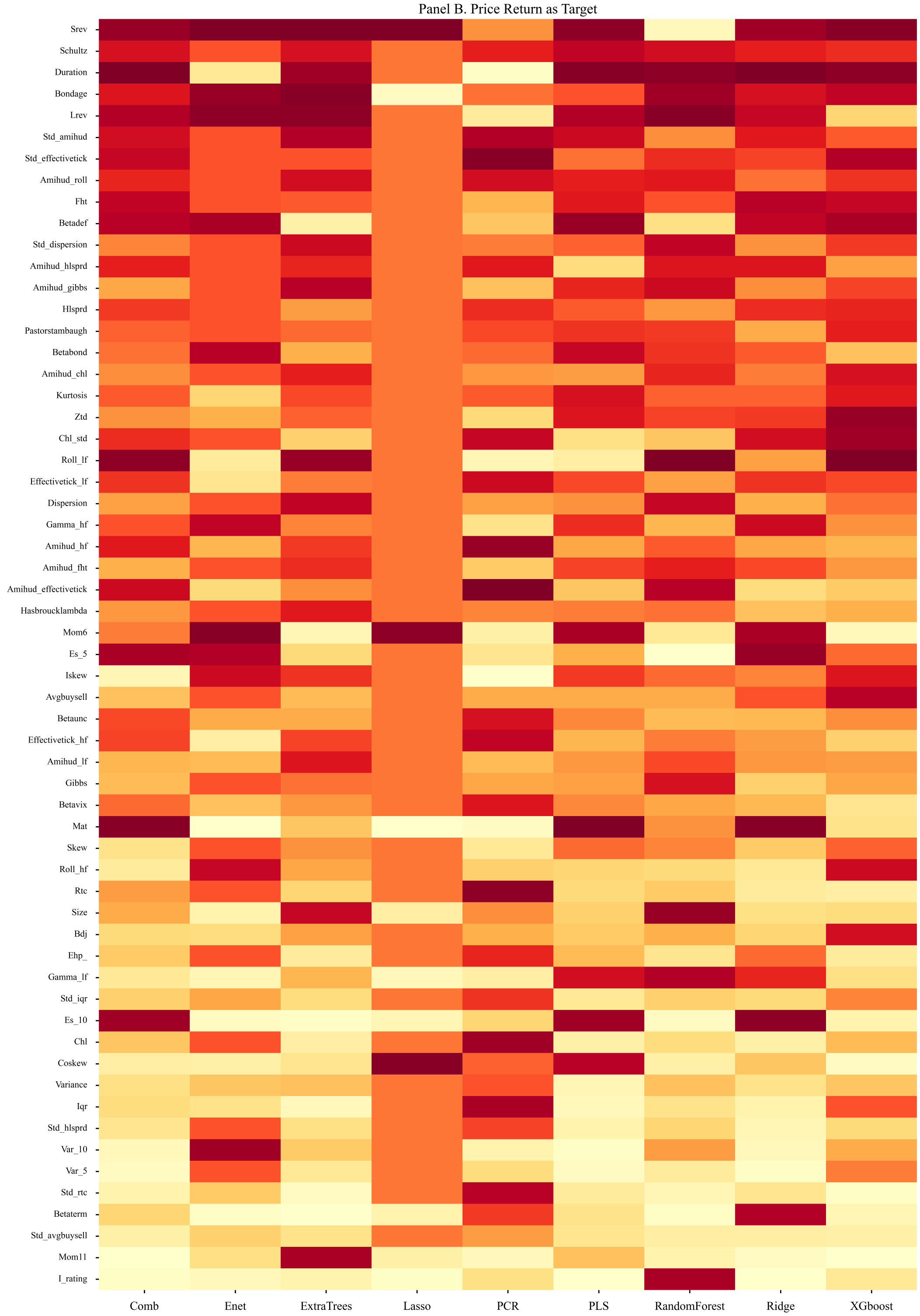


Figure 3: Portfolio-Level Out-of-Sample Prediction Performance for Public Firms

This figure displays the out-of-sample R^2 for 12 corporate bond portfolios using ordinary least squares (OLS), Lasso, ridge regression (Ridge), elastic net (ENet), principal component regression (PCR), partial least squares (PLS), XGBoost, random forest (RF), Extratrees , as well as the forecast combining method (Comb). Rows correspond to rating-maturity portfolios, which are formed at the beginning of each month based on ratings (AAA&AA, A, BBB, or speculative grades) and time to maturity (≤5 years, 5-10 years, >10 years) of individual bonds. Color gradients indicate the highest R^2_{OOS} values (dark orange) to the lowest ones (white). Results are based on 59 bond characteristics and 153 additional equity (firm-level) characteristics as bond return predictors among public firms. The testing sample spans from July 2004 to June 2019.

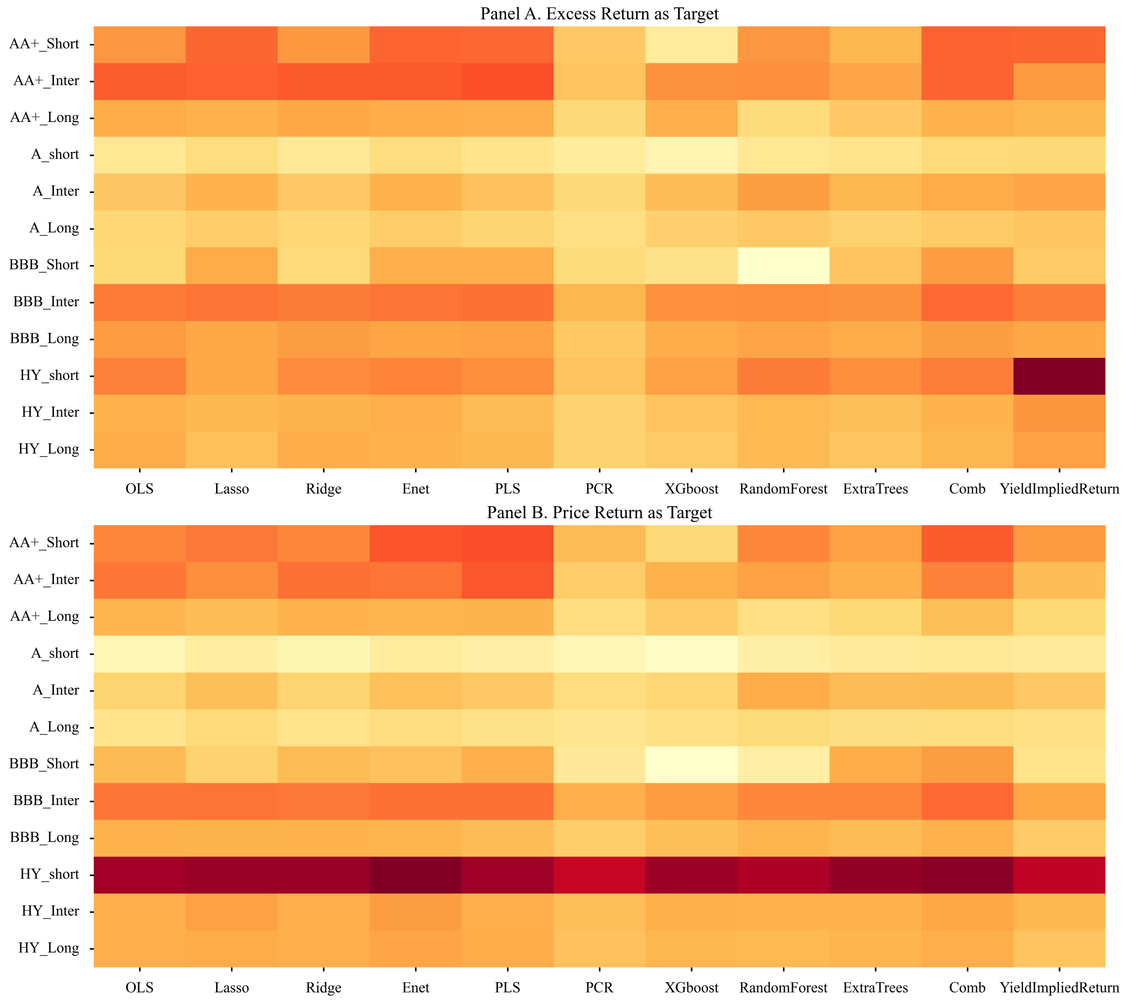


Figure 4: Portfolio-Level Out-of-Sample Prediction Performance for Private Firms

This figure displays the out-of-sample R^2 for 12 corporate bond portfolios using ordinary least squares (OLS), Lasso, ridge regression (Ridge), elastic net (ENet), principal component regression (PCR), partial least squares (PLS), XGBoost, random forest (RF), Extratrees , as well as the forecast combining method (Comb). Rows correspond to rating-maturity portfolios, which are formed at the beginning of each month based on ratings (AAA&AA, A, BBB, or speculative grades) and time to maturity ($\leq$5 years, 5-10 years, >10 years) of individual bonds. Color gradients indicate the highest R^2_{OOS} values (dark orange) to the lowest ones (white). Results are based on 59 bond characteristics as bond return predictors among private firms. The testing sample spans from July 2004 to June 2019.

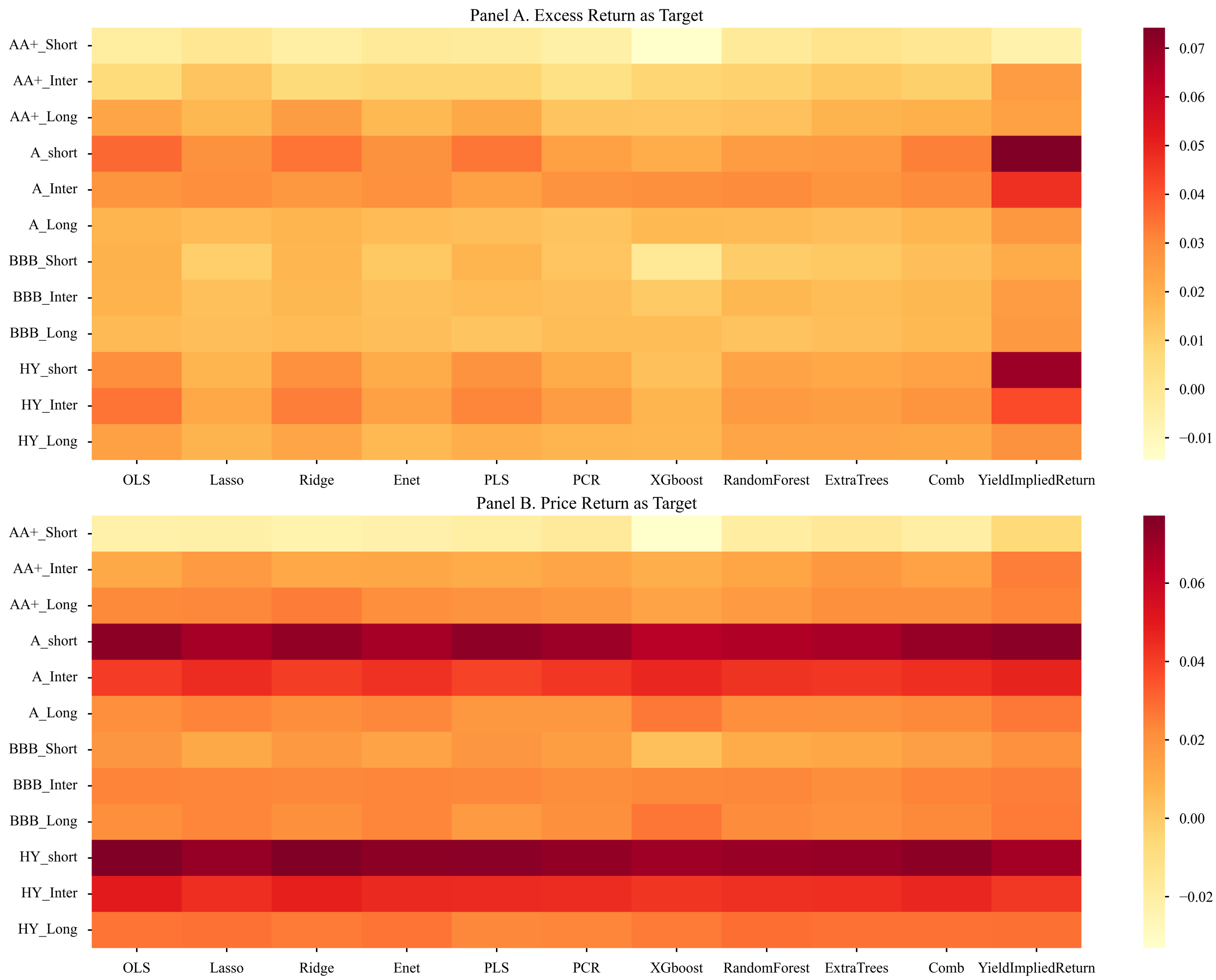


Figure 5: Market Timing Utility Gains from Machine Learning Forecasts for Public Firms

This figure displays the certainty equivalent gains (in percentage points) for a mean-variance investor who optimally invests in a corporate bond portfolio and a risk-free asset based on machine learning forecasts among public firms. Portfolios are formed at the beginning of each month based on ratings (AAA&AA, A, BBB, or speculative grades) and time to maturity ($\leq$5 years, 5-10 years, >10 years) of individual bonds. The weight on the risky bond portfolio is confined to lie between 0 and 150%.The investor's risk aversion coefficient γ is set at five. All machine learning models are benchmarked against the yield-implied expected bond returns. Color gradients differentiate negative certainty equivalent values (white) from positive ones (dark orange). Cells with asterisks denote statistical significance at the 1%, 5% and 10% level for an extended version of Diebold and Mariano (1995) test. Results in Panels A and B are based on excess return prediction and price return prediction, respectively. The sample period spans from July 2004 to June 2019.

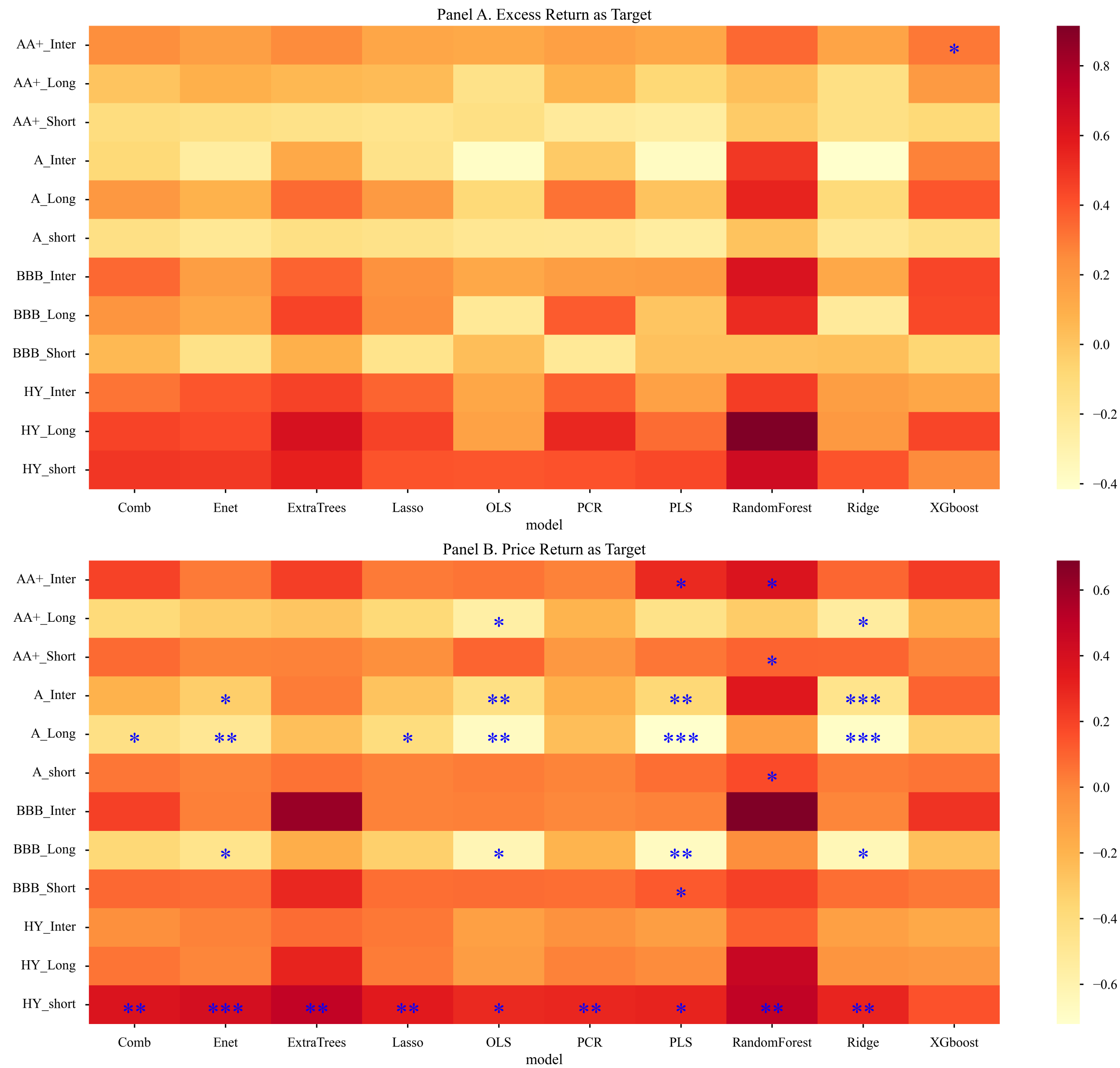


Figure 6: Market Timing Utility Gains from Machine Learning Forecasts for Private Firms

This figure displays the certainty equivalent gains (in percentage points) for a mean-variance investor who optimally invests in a corporate bond portfolio and a risk-free asset based on machine learning forecasts among private firms. Portfolios are formed at the beginning of each month based on ratings (AAA&AA, A, BBB, or speculative grades) and time to maturity (≤5 years, 5-10 years, >10 years) of individual bonds. The weight on the risky bond portfolio is confined to lie between 0 and 150%.The investor's risk aversion coefficient γ is set at five. All machine learning models are benchmarked against the yield-implied expected bond returns. Color gradients differentiate negative certainty equivalent values (white) from positive ones (dark orange). Cells with asterisks denote statistical significance at the 1%, 5% and 10% level for an extended version of Diebold and Mariano (1995) test. Results in Panels A and B are based on excess return prediction and price return prediction, respectively. The sample period spans from July 2004 to June 2019.

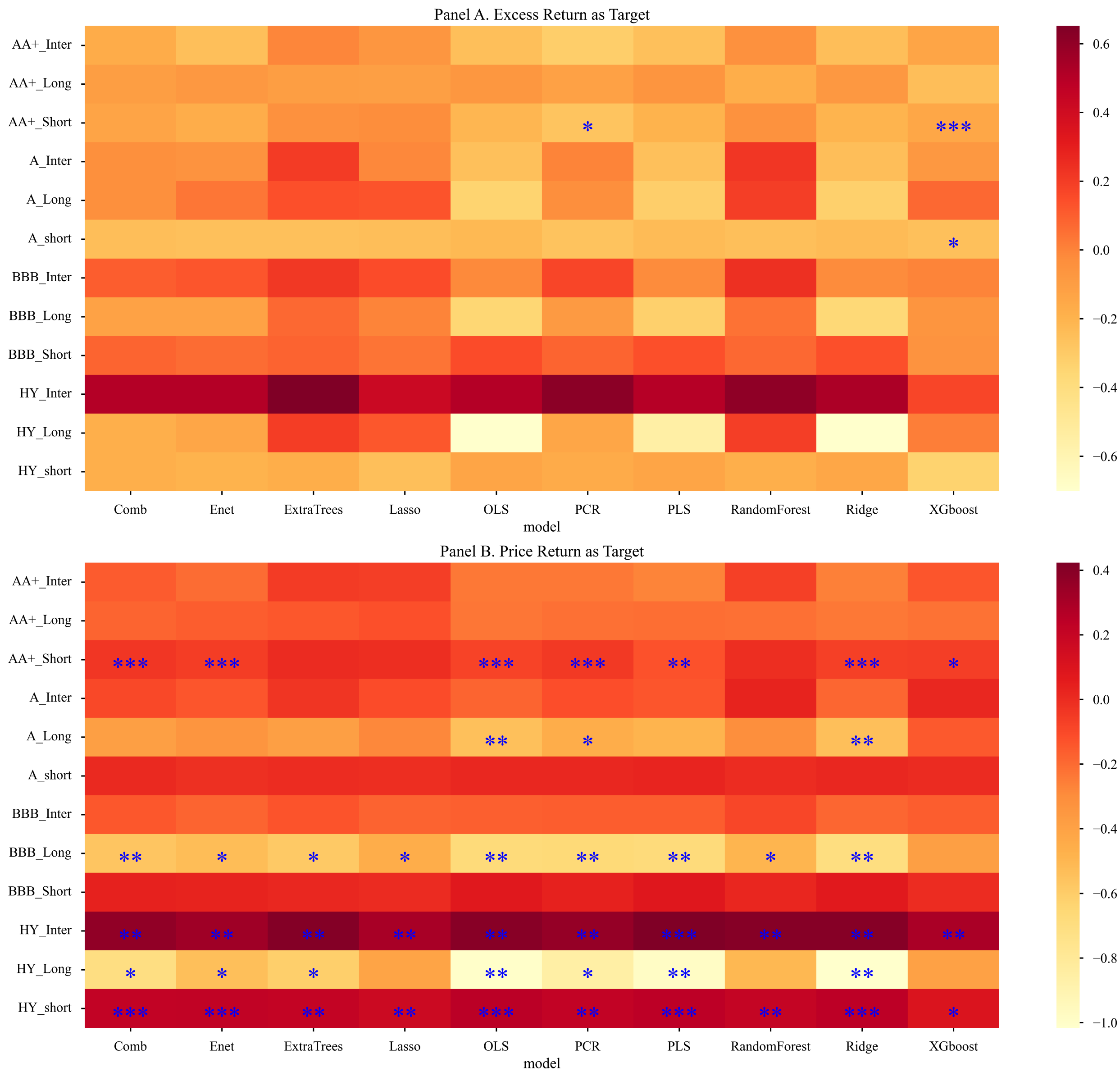


Figure 7: Investment Gains from Hedge Portfolios Based on Bond Return Forecasts for Public Firms

This figure presents the cumulative returns of long-short portfolios formed with different types of bond return forecasts. Decile sorted portfolios are constructed in each month for three prediction method—ordinary least squares (OLS), machine learning forecast combining method (Comb), and yield-implied expected bond returns (Yield). A long-short strategy is then built by longing the top return forecast portfolio and shorting the bottom return forecast portfolio. Panel A1 and B1 report the cumulative gross returns based on predicted excess returns and price returns, respectively. Panel A2 and B2 report the cumulative net returns after considering transaction costs based on predicted excess returns and price returns, respectively. The sample period spans from July 2004 to June 2019.

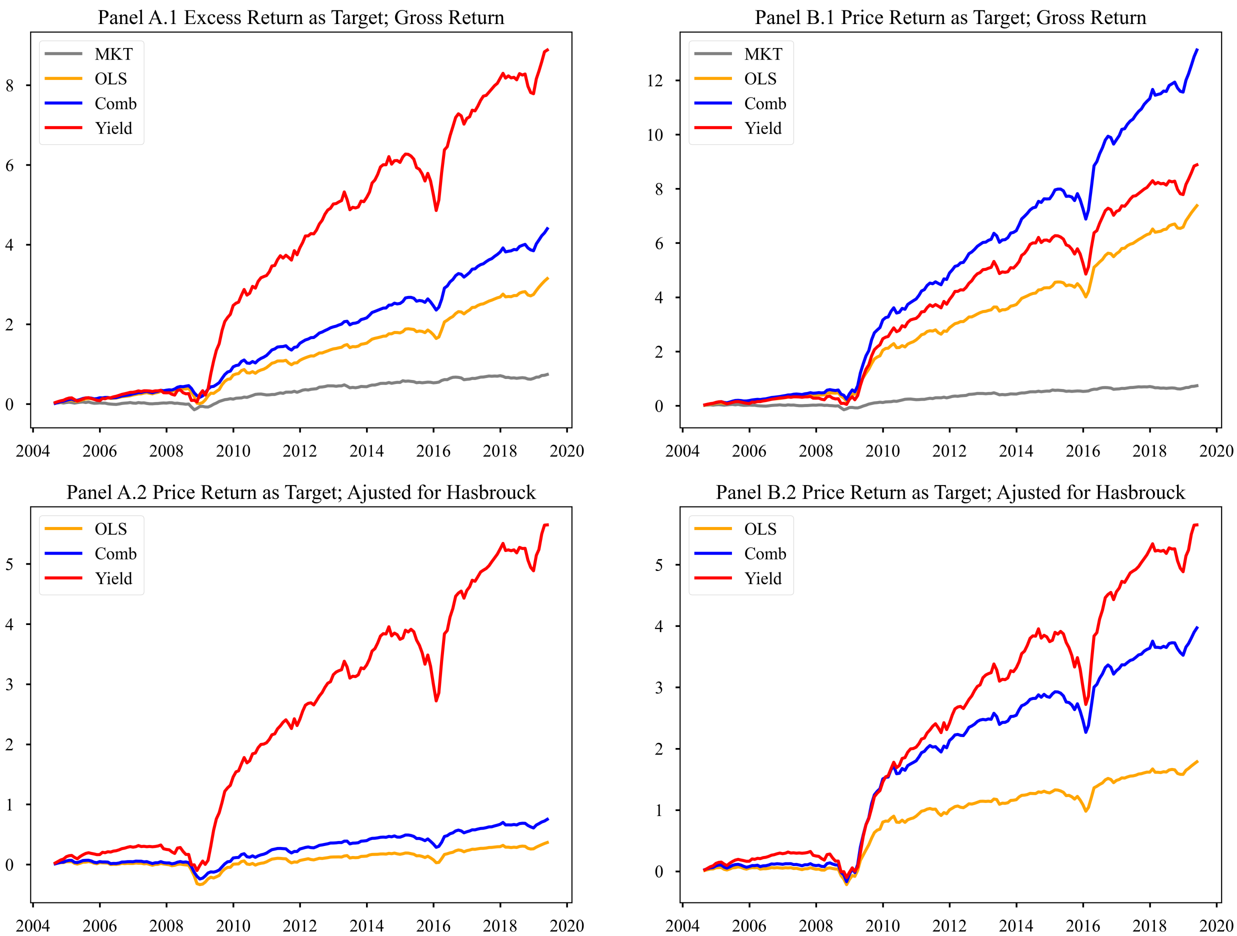


Figure 8: Investment Gains from Hedge Portfolios Based on Bond Return Forecasts for Private Firms

This figure presents the cumulative returns of long-short portfolios formed with different types of bond return forecasts. Decile sorted portfolios are constructed in each month for three prediction method—ordinary least squares (OLS), machine learning forecast combining method (Comb), and yield-implied expected bond returns (Yield). A long-short strategy is then built by longing the top return forecast portfolio and shorting the bottom return forecast portfolio. Panel A1 and B1 report the cumulative gross returns based on predicted excess returns and price returns, respectively. Panel A2 and B2 report the cumulative net returns after considering transaction costs based on predicted excess returns and price returns, respectively. The sample period spans from July 2004 to June 2019.