Benchmarking Individual Corporate Bonds *

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

First Draft: December, 2020

This Draft: August 6, 2021

Abstract

We propose an alternative approach to the linear factor model to estimate and decompose asset risk premia. To resolve the high-dimensional sort difficulty in forming characteristic-based benchmark portfolios, we introduce a *benchmark combination model (BCM)* that combines multiple basis portfolios as the pricing kernel. With a non-arbitrage objective, our approach minimizes cross-sectional pricing errors and identifies the sources of risk. We find that the BCM outperforms factor models in pricing corporate bonds. Credit ratings, downside risk and short-term reversal are primary sources of bond risk premia. Incorporating machine learning forecasts into the BCM, we find strong evidence of return predictability.

Key Words: Characteristic-based portfolios, high-dimensional sort, corporate bond risk premia, forecast combination, machine learning, return predictability.

*We appreciate insightful comments from Jianqing Fan, Dacheng Xiu, Li Guo, Xiaohu Wang and Zhonghao Fu. We are also grateful for helpful comments from seminar and conference participants at Fudan, Gothenburg, 2021 Financial Market and Corporate Governance Conference, 2021 China Accounting and Finance Conference, 37th International Conference of the French Finance Association, 2021 China International Risk Forum, and 2021 SoFiE Summer School at NYU Shanghai.

†College of Business, City University of Hong Kong. E-mail address: xin.he@my.cityu.edu.hk.

‡College of Business, City University of Hong Kong. E-mail address: gavin.feng@cityu.edu.hk.

§College of Business, City University of Hong Kong. E-mail address: jwang2@cityu.edu.hk.

¶School of Management, State University of New York at Buffalo. E-mail address: chunchiw@buffalo.edu.

1 Introduction

Asset pricing models can be used to evaluate whether an asset is underpriced or overpriced. Linear factor models, such as CAPM and Fama-French factor models, provide a benchmark for expected returns. An alternative approach uses the characteristic-based portfolios for inferring expected returns. Asset or industry characteristics are commonly used to construct portfolios as a benchmark to evaluate individual asset returns. Daniel et al. (1997) provide a characteristic benchmarks for equity and mutual fund returns based on portfolios formed by firm characteristics. They perform a three-way sort on size, value, and momentum and use the corresponding sorted portfolio as the asset pricing benchmark. For decades, practitioners such as Morningstar have been promoting the size-value equity style box and the rating-duration fixed-income style box as benchmarks for stocks and bonds.

Researchers usually classify individual corporate bonds into different credit rating categories (AAA/AA/A/BBB/Junk) to form portfolios. These are called basis portfolios, in which individual assets are sorted on one characteristic (ratings). Assets in the same basis portfolio are observed to bear similar credit risk and therefore, the corresponding rating-based portfolio serves as a performance benchmark for all individual corporate bonds with the same rating. In addition to credit ratings, researchers also consider maturity, duration, momentum, downside risk, and many bond characteristics to form basis portfolios.

An economic rational benchmark is supposed to include all information in these characteristics to reflect different risk exposures. However, there is a problem in high-dimensional sorts on multiple characteristics, as highlighted in Cochrane (2011). As an example, the number of corporate bonds in each sorted portfolio decreases as the number of characteristics increases. One may have portfolios with only a few or even zero observations due to the interactions among characteristics. The consequences include an increase for the estimation error and a high chance for an incorrectly model identification.

This paper proposes an alternative asset pricing evaluation framework, dubbed the benchmark combination model (BCM afterward), to alleviate the high-dimensional sort problem. Instead of sorting assets over multiple characteristics, we sort assets on each characteristic to construct basis portfolios. We then apply the linear combination approach on these basis portfolios associated with different characteristics to build the BCM evaluation. This method is transparent and objective. It is data-driven and preserves the rich information in all relevant characteristics. The linear modeling of high-dimensional characteristics also allows us to decompose returns and identify important sources of corporate bond risk premia. The BCM offers a asset pricing research paradigm that goes beyond corporate bonds, and can be easily applied to other asset classes, such as equities, commodities, and currencies.

In nutshell, the goal of our paper is fourfold: (1) we tackle the high-dimensional sort difficulty for cross-sectional returns; (2) we evaluate the model fitness with a non-arbitrage objective against conventional linear factor models and find our method exceptionally useful; (3) our model is intuitively interpretable and identifies the economic sources of corporate bond risk premia; and (4) we demonstrate the superiority of our model for predicting corporate bond returns.

In essence, BCM is equivalent to a non-parametric cross-sectional regression, where bond return observations are regressed on dummy variables of basis portfolios. Our method, a combination on univariate-sorted basis portfolios, is similar to the additive model of Cattaneo et al. (2020). In addition, Kelly et al. (2019) suggests that sorting on different characteristics provides ensemble samples to dissect the cross-sectional return distribution. The same bond being evaluated by different basis portfolios for different risk exposures can be viewed as the ensemble scheme. Each basis portfolio serves as a "rational" benchmark, and their linear combination helps reduce the bias and variance for expected returns. Most importantly, BCM is intuitively interpretable, where the common combination weights can be used to evaluate and decompose corporate bond risk premia.

Linear factor models are commonly used in empirical asset pricing because of their nice statistical properties and intuitive economic interpretation. However, one drawback of the factor model is its regression estimation accuracy relies on the sample size. For this reason, most empirical studies estimate factor models for portfolios instead of individual assets. BCM and factor model share the exact economic implementation purpose, benchmarking asset returns with a small number of trackable portfolios and decomposing risk premia. With less parametric modeling and hence less restrictive parametric assumptions, BCM stands a good chance to outperform conventional linear regression models on pricing individual asset returns.

We demonstrate the use of our model and importance of its results using a comprehensive data set that consists of 589,528 bond-month observations and covers 19,782 bonds from 1976 to 2017. We find that credit ratings, downside risk, and short-term return reversal are important sources of risk premia. In particular, short-term reversal is an under-appreciated pricing factor that shows high power. Our benchmark combination model outperforms conventional factor models in pricing corporate bond returns. In addition, there is strong evidence that the bond basis portfolio returns are predictable, and their combination predicts individual bond returns. The BCM carries high out-of-sample predictive power for individual bond returns with different machine learning forecasts. The forecast-implied long-short strategy delivers delivers a return of 11.88% per annum, which cannot be explained by the five-factor model of Fama and French (1993), and a annualized Sharpe ratio of 2.57. The results demonstrate the ability of the BCM to identify the relevant sources of risk embedded in characteristics to price corporate bonds and improve portfolio-sorting-based inference in asset pricing.

Our work is related to the literature on the pricing of individual corporate bonds based on the framework of risk factors and characteristics. Fama and French (1993) propose a five-factor model, and Bai et al. (2019) introduce common risk factors based on corporate bonds' risk characteristics, including downside risk, credit risk, liquidity risk, and short-term return reversal. Additionally, a few papers investigate the cross-section of corporate bond returns; see, for examples, Lin et al. (2011), Chung et al. (2019), Huang et al. (2020), and Gao et al. (2020). Recently, citekelly2020modeling studies individual corporate bonds using the instrumental PCA, which incorporates information of bond characteristics to estimate latent factors and factor loadings. This paper studies a similar problem and provides an interpretation of linear model combination.

This paper contributes to the literature on characteristic-sorted benchmarks. Daniel et al. (1997) provide a benchmark evaluation framework for individual equity returns and mutual fund performance, which dependently sorts on size, value, and momentum. We provide a solution to the high-dimensional sort difficulty highlighted in Cochrane (2011). Cattaneo et al. (2020) develop a general framework for portfolio sorting by casting it as a nonparametric estimator, and present valid asymptotic inference and methods for determining an optimal choice for the number of portfolios. Feng et al. (2020) develop a deep learning framework to approximate the characteristic-sorted factor model, while Feng et al. (2020) provide robust inference on the high-dimensional factor model.

Our paper is also linked to the literature of corporate bond return prediction. Hong et al. (2012) introduce nonlinear time series models, Lin et al. (2014) adopt combination forecasts, and Lin et al. (2018) develop an iterated combination approach. These papers show that corporate bond returns are predictable. Chordia et al. (2017) find find equity characteristics, such as profitability and asset growth, can predict corporate bond returns. Recently a number of papers investigate corporate bond return predictability using the machine learning approach (see Bredendiek et al. (2019), He et al. (2021) and Guo et al. (2020)) by exploiting information in bond or equity characteristics. Unlike these papers, this paper focuses on estimating and decomposing the sources of bond risk premia.

The remainder of the paper is organized as follows. Section 2 demonstrates the characteristic-sorting mechanism and the Benchmark Combination Model. The empirical research design is introduced in Section 3. Section 4 provides the empirical findings and illustrate the asset pricing performance of BCM. We also present return predictability evidence via BCM in Section 5. Section 6 summarizes the paper.

2 Methodology

2.1 Characteristic-Sorted Basis Portfolios

In this section, we describe the procedure to implement our model. The method starts with the formation of basis portfolios which are created by sorting assets on a single characteristic. The relationship between the excess return r and a single characteristic z lagged one period can be characterized by the following regression:

$$\begin{aligned} r_{i,t} &= \mu_t(z_{i,t-1}) + \epsilon_{i,t} && (1) \\ &= \gamma(z_{i,t-1}) + \nu_t(z_{i,t-1}) + \epsilon_{i,t}, && (2) \end{aligned}$$

for $i = 1, \cdots, n_t$ and $t = 1, \cdots, T$, and $\mu_t(\cdot)$ models the part of returns that can be explained by the characteristic-sorted basis portfolios, and $\gamma(\cdot)$ is the basis portfolio expected returns, and $\nu(\cdot)$ is the contemporaneous perturbation of basis portfolio return. Similar notations can be seen in Giglio and

Xiu (2021) for linear factor asset pricing models $r_t = \beta\gamma + \beta\nu_t + u_t$, where β is risk exposure, and γ is the factor risk premia.

We first sort individual corporate bonds on a single characteristic z and calculate the order statistics. With a given number of portfolios ($S = 5$), we can partition the cross-section into S number of bins. The first bucket contains the corporate bonds with the smallest z value, and the S-th bucket include those with the largest z value. The partition P is denoted as

$$P_{s,t} = \left[z_{(\lfloor n_t(s-1)/S \rfloor)t}, z_{(\lfloor n_t s/S \rfloor)t}\right], \tag{3}$$

for $s = 1, \cdots, S-1$, and

$$P_{S,t} = \left[z_{(\lfloor n_t(S-1)/S \rfloor)t}, z_{(n_t)t}\right], \tag{4}$$

where $z_{(l)t}$ is the order statistic, and $\lfloor \cdot \rfloor$ is the floor operator. Each characteristic $z_{i',t}$ corresponds to one partition, and we can write the partition index as a function of characteristic $s' = s(z_{i',t})$. The partition index s' implies the basis portfolio return $R_{s',t}$,[1]

$$\hat{\mu}_t\left(z_{i',t-1}\right) = R_{s',t} = \frac{1}{|P_{s't}|} \sum_{i:z_{i,t-1} \in P_{s't}} r_{i,t}, \tag{5}$$

where $|P_{s't}|$ denotes the number of bond observations in partition $P_{s't}$.

As in Cattaneo et al. (2020), calculating the equally-weighted basis portfolio return is equivalent to estimating a cross-sectional regression. We can construct a dummy variable to record the sorting index s, $\hat{\mathbf{1}}_{s,t}(z) = \mathbf{1}\{z \in P_{s,t}\}$. Hence, we can estimate $\gamma(\cdot)$ by regressing individual bond returns on dummy variables.

2.2 High-dimensional Sort Difficulty

Extending from a univariate-sort on one characteristic to a multivariate-sort on multiple characteristics is more complicated. It depends on the high-dimensional sorting setting, such as a dependent sort or an independent sort. In case of independent sort, suppose the number of partitions is S for each characteristic and K is the number of characteristics. There are a S^K number of parti-

[1] For demonstration, we use equally-weighted portfolio return. It is easy to extend to value-weighted portfolio return.

tions, and the average number of corporate bonds in each partition is $n/(S^K)$. For a sample of 3,000 bonds, with 5 partitions, and 5 characteristics, the average portfolio size is 4.8. Researchers hope to form a portfolio containing a sufficient number of corporate bonds to represent the common risk exposure and diversify the idiosyncratic risk. However, high-dimensional sorting shrinks the size of each portfolio dramatically. It is likely that some partition contains even fewer corporate bonds under independent sorting, especially when characteristics are highly correlated.

Table 1: The Number of Observations by Bins for Bivariate Sorts

This table reports the monthly average number of observations in each bucket for independent bivariate sorts from 2015 to 2017. We sort individual corporate bonds by credit rating, duration, and downside risk into five bins. The top table shows the number of observations for rating and duration sorts, and the bottom table shows the bivariate sorts on duration and downside risk.

		Duration					
		1	2	3	4	5	All
Rating	AAA	42	39	32	16	30	158
	AA	38	30	19	15	23	125
	A	114	104	81	92	104	494
	BBB	78	86	106	141	122	532
	Junk	22	34	55	31	18	161
	All	293	293	293	295	296	1471
Downside risk	1	9	15	69	48	156	296
	2	15	30	71	74	105	295
	3	42	59	79	83	33	296
	4	73	62	65	90	2	292
	5	154	127	9	1	1	292
	All	293	293	293	295	296	1471

As an example, in Table 1, we report the bivariate sort of corporate bonds on duration and rating in the upper panel and duration and downside Risk" in the lower panel. There are a few bins with less than ten observations. Implementing a high-dimensional sort is difficult as indicated by Cochrane (2011). Though the benchmark evaluation through this sorting mechanism is straightforward, it can be quite challenging when there are more than two characteristics.

2.3 The Benchmark Combination Model

The BCM we propose is a combination approach which has distinct advantages when facing a large number of characteristics in portfolio sorting. Each basis portfolio can be useful as a benchmark for asset pricing, and combining them reduces the potential bias and variance. With K characteristics, in matrix form $Z_{i,t-1}$, and thus K basis portfolios, we denote the basis portfolio on characteristic k as $R_{k,s,t}$, where s is partition index, and t is time index. We combine K basis portfolios into one benchmark $\mu_t(Z_{i,t-1})$ with the combination weight $\omega = [\omega_1, \omega_2, \cdots, \omega_K]^\intercal$:

$$\mu_t(Z_{i,t-1}) = \sum_{k=1}^{K} \omega_k R_{k,s,t}, \tag{6}$$

where $R_{k,s,t}$ is the basis portfolio return of asset i based on characteristic k, and the notation for asset i is embedded in the basis portfolio index s. Now, the sorting function s is generalized to $s = s_k(\cdot)$, for different k. For example, if a corporate bond has an AAA rating with a duration of 30 years and median downside risk, then the partition index s is 1 for the rating, 5 for duration, and 3 for downside risk. Therefore, the same corporate bond i can be assigned to different s, depending on the characteristic k of interest.

2.4 BCM Estimation

The estimation procedure of BCM weight follows the arbitrage pricing theory. Similar to Feng et al. (2020), we formulate the estimation as an optimization problem to minimize the pricing error, and implement a (constrained) panel regression. To reserve the economic interpretability, we add non-negativity and sum-to-unity constraints on common combination weights:

$$\underset{\{\omega_1,\omega_2,\cdots,\omega_K\}}{\text{Minimize}} \ \mathcal{L} = \sum_{i,t} (r_{i,t} - \sum_{k=1}^{K} \omega_k R_{k,s,t})^2 \tag{7}$$

$$\text{with} \sum_{k=1}^{K} \omega_k = 1, \omega_k \geq 0$$

The constrained BCM is denoted as BCMc in empirical investigation. We also provide an unconstrained version, BCMu. The panel regression helps to identify important sources of risk when decomposing the risk premia for individual corporate bond returns.

2.5 Discussion: Compare with Factor Model

The BCM provides a linear pricing kernel for individual corporate bond returns. A synthetic portfolio is implied for the individual corporate bonds, which allocates $\{w_k\}$ weights on basis portfolios $\{R_{k,s,t}\}$. The conventional factor models have a similar pricing kernel interpretation. For example, the CAPM can be interpreted through the benchmark $\mu_t(Z_{i,t-1})$ notation. The pricing kernel for asset i is the product of two components: the factor loading β_i and the market factor f_t,

$$\beta_i \times f_t. \tag{8}$$

The CAPM provides a synthetic portfolio of individual asset returns, which allocates β_i weight to the market factor. The synthetic portfolio captures the systematic risk of an individual asset, and provides an economic interpretable benchmark for individual asset evaluation.

The estimation and interpretation for the BCM are similar to the multifactor model under the non-arbitrage condition in linear modeling. However, the multifactor model estimates factor loadings for different assets ($N \times K$ parameters for a K-factor model to N assets). By contrast, our BCM uses all assets to estimate the common combination weights (K parameters for K characteristics). The partitions give specific pricing properties for basis portfolios. By comparison, BCM has a larger estimation sample with a much fewer number of arameters.

2.6 BCM Return Prediction

With the pricing kernel, an immediate application is to perform return prediction. To demonstrate the usefulness of the model, we first replicate the results of predictability in Lin et al. (2014) and Lin et al. (2018) on rating or duration basis portfolio returns. We then extend their analysis by presenting return predictability evidence for various basis portfolios. Using the BCM, we show that portfolio return predictability can be transformed into individual bond return predictability. The conditional forecast of individual bond returns is straightforward:

$$E_t(r_{i,t+1}) = E_t\big(\mu_{t+1}(Z_{i,t})\big) = E_t(\gamma(Z_{i,t})) = \sum_{k=1}^{K} \omega_{k|t} E_t(R_{k,s,t+1}), \tag{9}$$

where $E_t(r_{i,t+1})$ is the conditional expectation of asset i's return at time $t+1$ based on information up to time t, $w_{j|t}$ is the combination weight estimated at time t, and $E_t(R_{k,s,t+1})$ is the conditional expectation of basis portfolio returns.

$E_t(R_{k,s,t+1})$ is essentially a prediction for the basis portfolio return. We formulate it in a general predictive function $g(\cdot)$:

$$E_t(R_{k,s,t+1}) = g_{k,s|t}(\widetilde{Z}_{k,s,t}, x_t), \tag{10}$$

where two groups of predictors are included: $\widetilde{Z}_{k,s,t}$ are aggregated characteristics for different basis portfolios s of characteristic k, and x_t are macro predictors up to time t.[2] We plug the forecast $E_t(R_{k,s,t+1})$ into Equation 9 to predict individual corporate bond returns.

3 Empirical Design

3.1 Performance Measures

3.1.1 Pricing Measure

Once the weight is estimated, we use Equation 6 to construct our model-implied return benchmark for each individual corporate bond. Given the targets are individual asset returns, we adopt the pricing performance measures total R^2 and predictive R^2 introduced in Kelly et al. (2019). Both measures can be used in the in-sample and out-of-sample studies.

$$\text{Total } R^2 = 1 - \frac{\sum_{i,t}\left(r_{i,t} - \sum_{k=1}^{K}\hat{\omega}_k R_{k,s,t}\right)^2}{\sum_{i,t} r_{i,t}^2} \tag{11}$$

where $\hat{\omega}_k$'s are the estimates of weights. The total R^2 represents the fraction of the realized return variation explained by the expected returns $\gamma(\cdot)$ and the contemporaneous basis portfolio returns perturbation $\nu_t(\cdot)$, aggregated over all assets and all periods. For in-sample pricing exercises, $\hat{\omega}_k$'s are estimated with the whole sample of interest, and the total R^2 is calculated. On the other hand, for out-of-sample pricing exercises, the weights are estimated in a rolling-window manner with the available information up to time $t-1$, denoted as $\hat{\omega}_k = \hat{\omega}_{k|t-1}$.

We calculate the predictive R^2 as follows:

$$\text{Predictive } R^2 = 1 - \frac{\sum_{i,t}\left(r_{i,t} - \sum_{k=1}^{K}\hat{\omega}_k \bar{R}_{k,s,t-1}\right)^2}{\sum_{i,t} r_{i,t}^2} \tag{12}$$

[2]The details of prediction implementation follow He et al. (2021).

where $\bar{R}_{k,s,t-1}$ denotes the 20-year moving average of the basis portfolio returns up to time $t-1$ for risk premia estimates. The predictive R^2 represents the fraction of the realized return variation explained by the model-implied expected returns $\gamma(\cdot)$, excluding the $\nu_t(\cdot)$ part. For comparison with factor models, we use the same procedure for in-sample and out-of-sample estimation and calculate similar performance measures suggested by Kelly et al. (2019).

3.1.2 Prediction Measure

To evaluate the predictive power of the BCM, we report the out-of-sample R^2 for predicting basis portfolio returns. We calculate the prediction error for basis portfolio return $\hat{R}_{k,s,t}$ by comparing with its 20-year moving average $\bar{R}_{k,s,t}$.

$$R^2_{OOS,k,s} = 1 - \frac{\sum_t (R_{k,s,t} - \hat{R}_{k,s,t})^2}{\sum_t (R_{k,s,t} - \bar{R}_{k,s,t})^2}, \quad \text{for } k \text{ in } 1, \cdots, K, \text{ and } s \text{ in } 1, \cdots, S. \tag{13}$$

We aggregate the prediction errors across portfolios sorted on each characteristic k.

$$R^2_{OOS,k} = 1 - \frac{\sum_{s=1}^{S=5} \sum_t (R_{k,s,t} - \hat{R}_{k,s,t})^2}{\sum_{s=1}^{S=5} \sum_t (R_{k,s,t} - \bar{R}_{k,s,t})^2}, \quad \text{for } k \text{ in } 1, \cdots, K. \tag{14}$$

We evaluate the predictability of individual bond returns with the out-of-sample R^2 below:

$$R^2_{OOS} = 1 - \frac{\sum_{i,t} (r_{i,t} - \hat{\mathcal{X}}_{i,t})^2}{\sum_{i,t} (r_{i,t} - \bar{r}_{i,t})^2}. \tag{15}$$

where the baseline prediction $\bar{r}_{i,t}$ is the 20-year moving average return for the corresponding rating basis portfolio.[3]

3.2 Fama-MacBeth Tests

The above total, predictive, and out-of-sample R^2's are aggregate measures over all time periods. However, a drawback for such aggregate measures is they cannot reflect the performance variation over time and can be dominated by extremely volatile periods. To overcome this problem, We follow He et al. (2021) and adopt the Fama-MacBeth measures for pricing and prediction.

[3] For example, a Rating-1 bond uses the average returns of the Rating-1 portfolio as the $\bar{r}_{i,t}$ in denominator.

In particular, we utilize a large cross-section of bonds and aggregate the information for each time period to obtain the R_t^2. Therefore, we are able to perform the Fama-MacBeth test on the average performance for the time series $\{R_t^2\}_{t=1}^T$ with the following hypothesis:

$$\text{H0: } \overline{R^2} > 0 \text{ ; H1: } \overline{R^2} \leq 0.$$

For example, we can calculate the periodical version of total R_t^2 as

$$\text{Total } R_t^2 = 1 - \frac{\sum_i \left(r_{i,t} - \sum_{k=1}^K \hat{\omega}_k R_{k,s,t} \right)^2}{\sum_i r_{i,t}^2}. \tag{16}$$

Given the flexibility of the Fama-Macbeth test design, we can apply the "two-sample t-test" for performance comparison,

$$\text{H0: } \overline{R_\Delta^2} = 0 \text{ ; H1: } \overline{R_\Delta^2} \neq 0,$$

where $R_{\Delta,t}^2 = R_{1,t}^2 - R_{2,t}^2$. We can perform this comparison test on two different methods or two different samples.

3.3 Implementation procedure

Implementing our method involves two steps. First, we need to estimate the basis portfolio weights for BCM. To do this, we perform the constrained regression for realized returns in Equation 7. For out-of-sample estimation, we update the model annually using the rolling window of the past 60 months. To obtain the significance of portfolio weights, we employ 1,000 bootstrap samples and report the non-negative estimation frequency for each benchmark to measure the significance of weights. Variables estimated with zero coefficients are useless for the objective in the bootstrap sample. A variable that appears more than 90% of the time is deemed significant. We present the bootstrap significance levels by dark and light color in Table 4.

Second, we build the return forecasts for each basis portfolio of all characteristics. Specifically, we train the prediction model by the time-series modeling instead of the pooled modeling. For example, the predictive model for rating-1 portfolio only uses its own data and does not involve data for other rating portfolios. The out-of-sample period is January 1998 to December 2017, and the training window is the rolling window of the past 20 years. We update the model annually and predict all 12 monthly observations for the following year. The prediction design follows He et al. (2021). Besides the moving average, we consider four major forecasting methods, including the mean combination, principal component regression, Lasso, and Random Forest.

4 Empirical Results

4.1 Data

The corporate bond data are from four major sources: the Lehman Brother Fixed Income (LBFI) database, DataStream, the National Association of Insurance Commissioners (NAIC) database, and the Enchanced Trade Reporting and Compliance Engine (TRACE) database. We combine the data from these sources to generate a large individual corporate bond sample. When there are duplicate observations from multiple sources, we keep only one of them by a priority rank. The rank from high to low priority is TRACE, NAIC, LBFI, and DataStream. Whenever there is a choice, we prefer the transaction-based return data (TRACE) to the return data based on quotes and matrix calculations (LBFI). The enhanced TRACE data start in 2002. Using the LBFI and NAIC data extends our sample to early 1973.

We calculate corporate bond returns using the combined sample. The monthly corporate bond return at time t is calculated as follows:

$$R_t = \frac{(P_t + A_t) + C_t - (P_{t-1} + A_{t-1})}{P_{t-1} + A_{t-1}}, \tag{17}$$

where P_t is the price, A_t is the accrued interest, and C_t is the coupon payment, if available, at time t. We obtain excess returns by subtracting the three-month Treasury bill rate from the raw return. Our sample excludes bonds with embedded options and bonds with maturity less than two years or longer than 30 years. Summary statistics of the bond sample are shown in Table 2.

[Insert Table 2 here]

We consider a comprehensive list of predictors, including 20 macroeconomic predictors and

20 corporate bond characteristics.[4] Detailed descriptions on the predictors are included in Table 3. Lin et al. (2014) and Lin et al. (2018) find that macroeconomic variables contains rich information for future corporate bond returns. Our macro predictor set covers two main categories: corporate bond market variables (e.g., Treasury bill rate, rating spread), and equity market variable (e.g., S&P 500 index returns, S&P 500 index earnings-to-price ratio). Many of the latter variables are used also by Welch and Goyal (2008).

The corporate bond characteristic sample covers three categories: fundamental characteristics (e.g., ratings, duration), return-distribution characteristics (e.g., momentum, downside risk), and covariances with common risk factors (e.g., beta of term and default spreads). The sample period runs from January 1973 through December 2017. Because we need three years of data to calculate the return-based characteristics, the full sample of predictors starts from January 1976. The out-of-sample evaluation period is from January 1998 to December 2017.

[Insert Table 3 here]

Based on each characteristic, we construct the basis portfolios (univariate-sorted quintile portfolios) and rebalance them every month. We calculate the basis portfolio returns as the equally-weighted average of the underlying assets' returns. Also, the characteristics of the basis portfolios are the equally-weighted average of the underlying assets' characteristics. The basis portfolio characteristics serve as the individual (idiosyncratic) signal in Equation 10.

4.2 Bond Risk Premium Decomposition

As discussed in Section 2, the BCM is an alternative approach to the factor model to provide an expected return and decompose risk premia. The first column in Panel A of Table 4 reports the in-sample results, using all return observations from 1998 to 2017. We regress individual bond realized returns on their corresponding basis portfolios. In Panel A for the constrained model, we find credit rating, short-term return reversal, 12-month momentum, and downside risk are important sources of bond risk by bootstrap inference. Under different economic conditions, the basis portfolio importance can be different in the in-sample analysis. We split the 20 years from 1998 to 2017

[4] The corporate bond characteristics and macro predictors are available to all corporate bonds, and we don't consider any equity characteristics, because they are only applicable to the corporate bonds issued by publicly listed firms.

into the periods of economic boom and recession using the NBER business-cycle indicators.[5] Credit rating is persistently strong, but duration and short-term reversal are more important in the economic boom than in the recession. Also, downside risk and 12-month momentum are significantly pricing sources in the recession but not in the period of economic boom.

The remaining columns report the out-of-sample results of common combination weights. The combination weights in each year are estimated using the past five-year rolling-window sample. First of all, most combination weights are zero due to the constraint estimation. Second, short-term Reversal is persistently strong in the entire sample. Third, credit ratings become dominant since the year of 2002. Finally, the downside risk is only important right after the 2008-2009 financial crisis.

In Panel B of Table 4, we report the unconstrained common combination weights. It might be difficult to interpret the negative weights in the economic context, though the main drivers (credit rating, short-term reversal, 12-month momentum, and downside risk) stand out with positive weights and significant t-stats. The unconstrained model serve as a comparison to the constrained model in empirical results. It is possible to improve the results by removing the constraints.

[Insert Table 4 here]

4.3 Asset Pricing Model Performance

4.3.1 Pricing Performance

Table 5 show the pricing performance of two BCM and two factor models. BCMc is the constrained model, and BCMu is the unconstrained one. The factor models follow the five-factor model (MKT, SMB, HML, TERM, DEF) in Fama and French (1993), and the four-factor model (CBMKT, CRF, DRF, STR) in Bai et al. (2019).[6]

Panel A reports in-sample total R^2, where BCM common combination weights are estimated from 1998 to 2017, and the individual bond factor loadings are estimated with the entire history of each bond.[7] Panel B reports out-of-sample total R^2, where BCM combination weights are dynami-

[5] See https://www.nber.org/research/business-cycle-dating.

[6] We follow Bai et al. (2019) to construct corporate bond market (CBMKT), downside risk (DRF), credit rating (CRF), and short-term reversal (STR) factors using the data in our sample.

[7] Given high volatility on individual bond betas, we have winsorized the beta estimates by 5% and 95% quantiles in the cross-section.

cally updated with the past five-year data.

Individual bond factor loadings have huge estimation errors. We can only find a similar in-sample total R^2 between BCM and factor models using the winsorized beta estimates for the factor model. The BCM is more robust than the factor model for pricing individual assets with limited observations. The BCM significantly outperforms factor models in terms of in-sample Fama-Macbeth $\overline{R^2}$, which shows the BCM performance is more stable over time.

For out-of-sample Total R^2, as shown in Panel B of Table 5, the BCM substantially outperforms the factor models, which even underperform the zero return baseline. We present the predictive R^2 in Panel C. With pre-estimated combination weights and basis portfolio return forecasts, the predictive R^2 describes the individual bond risk premia with a pure forecasting perspective. Negative numbers in predictive R^2 imply that a model underperforms the zero average return baseline. We find both BCM frameworks generate significantly positive predictive R^2, while the factor models generate insignificant or even negative R^2's. Finally, these results are robust for investment-grade and non-investment-grade bonds and suggest that the BCM is a better alternative to estimate the risk premia for individual bonds.

[Insert Table 5 here]

4.3.2 Model Comparison

Given the large cross-section of the panel data, the Fama-MacBeth test can be very useful for multiple comparisons. We consider a two-sample Fama-MacBeth test to compare the $\overline{R^2}$ between the BCM and factor models. In each panel of Table 5, we also report the performance difference in R^2, and its average $\overline{R^2_\Delta}$, of the BCM versus factor models, for example "BCMc-FF5". Both BCM models deliver significantly larger $\overline{R^2_\Delta}$ than the FF5 and BBW4 models in all pricing metrics and different subsamples.

Since the bivariate-sorted benchmark on rating and duration are commonly used, we also provide an example to compare the BCM with this set of benchmark portfolio to evaluate the additional value of using multiple characteristics. The predictive R^2's are reported in Table 6. In Panel A, we find the BCMc model produces significantly positive numbers for long-duration (and high-rating) bonds. We also find the positive performance of moving-average returns for the bivariate-sorted benchmark over the zero average return baseline. Finally, Panel C shows that the BCMc outperforms the moving-average model for bivariate-sorted portfolios.

[Insert Table 6 here]

4.3.3 Private versus Public Bonds

Past studies (Choi and Kim, 2018 and Chordia et al., 2017) show that equity characteristics contain information for corporate bond returns in the cross section. This analysis is only feasible for those corporate bonds issued by publicly listed firms that also issue stocks. For bonds issued by private firms, access to equity characteristics is limited or even impossible. The above studies only consider corporate bonds issued by publicly listed firms in their sample. The usefulness of equity characteristics can be due to the sample selection for publicly listed firms. In our bond sample more than 75% of bond-return observations are from private firms. (also see Table 2). This raises an issue of whether the finding of equity characteristic predictive power is robust to private bonds. There is a concern that past findings could be due to sample selection.

We approach this issue differently by examining whether public and private bonds have the same pricing difficulty (signal-to-noise ratio) when using the same predictors without equity characteristics. If the corporate bond market is efficient, the relative pricing performance for public and private bonds should equal. If we find similar a pricing performance for public and private bonds without using any equity characteristics, then equity characteristics may be redundant information.

Panels A and B of Table 7 report the total R^2 for private bonds, public bonds, and their difference, whereas Panel C reports the predictive R^2. For the investment-grade (IG) and whole samples, We find private bonds have significantly smaller total R^2 and predictive R^2 than public bonds based on the Fama-Macbeth tests. For the non-investment-grade (NIG) subsample, although we find the same sign of difference, it is not statistically significant at the 10% level. Without using any equity information, we find the public bonds have smaller pricing errors than private bonds, and this finding concentrates on the IG bonds. For the subperiod analysis in the rest of the panels, we find the performance differences between private public bonds appear mostly during and after the 2008 global financial crisis. Overall, there is no convincing evidence that equity characteristics are redundant for pricing public bonds.

[Insert Table 7 here]

5 Predicting Returns Using the BCM

The results above suggest that the BCM serves very well as a pricing kernel. An issue of considerable interest is whether we can use the BCM for predicting future bond returns. The BCM return forecast must be built on the basis portfolio return forecast as a combination of basis portfolios. In this section, We provide evidence of return predictability using both conventional and machine learning methods. We find that the basis portfolio return forecasts are helpful for predicting the underlying individual asset returns. Given the basis portfolio return forecasts, we can calculate the BCM return forecast for individual assets. We show that trading strategies using the BCM return forecasts generate positive investment performance.

5.1 Basis Portfolio Return Predictability

We first present the return predictability evidence for basis portfolios. The aggregate out-of-sample R^2 (Equation 14) of five basis portfolios for each characteristic are reported. Consistent with Lin et al. (2014) and Lin et al. (2018), we find that mean combination can predict the rating or duration basis portfolio returns. Furthermore, we find that machine learning methods can substantially improve forecast performance, implying a potential nonlinear predictive relationship between the predictors and bond returns.

In Table 8, we extend the return predictability studies to all 20 groups of basis portfolio returns. For all basis portfolios, we find high return predictability over the average benchmark. The traditional mean combination forecast method delivers robust predictability for all basis portfolios. Moreover, machine learning methods, including Lasso, PCA regression, and Random Forest, generate strong return predictability. The results constitutes one important contribution for this paper: basis portfolio returns are substantially predictable, especially via machine learning methods.

[Insert Table 8 here]

5.2 Predicting Individual Bond Returns by Basis Portfolio Prediction

Before demonstrating the BCM predictive power, we show the predictability for single basis portfolios on individual bond returns. Table 9 reports the out-of-sample R^2_{OOS} of predicting individual bond returns (see Equation 15) for different prediction models: average return prediction, mean combination forecast, LASSO, PCA, and Random Forest. For computing the denominator in Equation 15, we adopt the baseline model $\bar{r}_{i,t}$ as average returns of the credit rating basis portfolio. Therefore, the values are zero for the row of "Rating" in the first panel "Average".

Mean Combination, Lasso, PCA regression, and Random Forest offer a substantial improvement over the average baseline forecasts. In particular, the Random Forest forecasts show highly significant return predictability improvements (in FM-$\overline{R^2}$) to each basis portfolio for different bond characteristics. Table 8 shows that the machine learning forecasts for basis portfolio returns are much more powerful than the average return forecast. These results are consistent with the findings in He et al. (2021), who show the power of machine learning in forecasting individual corporate bond returns. Table 9 further shows that basis portfolio returns can be used to predict individual corporate bond returns. These are the empirical foundation for a combination forecast on basis portfolios to individual bond returns. On the top panel of Table 9, we also present the constrained and unconstrained BCM forecasts, which outperform most individual basis portfolio forecasts.

[Insert Table 9 here]

5.3 Predicting Individual Bond Returns by the BCM

We next demonstrate how to generate return forecast using the benchmark combination model's pricing kernel. In step one, we follow the previous subsection and create the return forecast for BCM components and various basis portfolios. In step two, we combine these return forecasts with BCM common combination weights to obtain a prediction model. Lastly, with the dynamic weights in Table 4, we construct a combination prediction (or ensemble prediction) for individual bond returns.

Panel A of Table 10 reports the out-of-sample performance of the benchmark combination forecast. Taking Lasso as an example, we have Lasso return predictions for various basis portfolios and combine these Lasso return forecasts with BCM weights. We then report the out-of-sample

R^2_{OOS} of predicting individual bond returns by the weighted combination Lasso forecast, labeled "c-Lasso" in the table. We find that traditional mean combination delivers slightly positive results, while Lasso, PCA regression, and Random Forest show substantial out-of-sample predictability.

Panel B reports the BCM forecast with unconstrained weights for a robustness check (labeled "u-Model" in the table). In Panel C, we find that the equally-weighted BCM provides positive results (labeled "m-Model" in the table.), but the magnitude is smaller than the optimally-weighted BCM results. We also report the results for the factor model, though most numbers are negative. The rest of the panels in Table 10 report the subperiod prediction results for individual bond returns. We find the predictability is consistently positive for all subperiods but stronger during and after the 2008-2009 financial crisis.

[Insert Table 10 here]

5.4 Investment Performance

Based on BCM forecasts for individual corporate bond returns, we construct forecast-implied long-short portfolios. Table 11 reports the performance measures for these long-short portfolios. We report the average returns, α's based on the five-factor model (MKT, SMB, HML, TERM, DEF), t-stat's for the alpha, and the annualized Sharpe ratio. We find that the average returns are substantially larger than the corporate bond market portfolio excess returns for three different samples (investment-grade, non-investment-grade, and overall bonds).

These positive investment performances are robust for all listed methods, including the BCM forecasts using basis portfolio average returns. The α's are positive and close to the average returns, indicating the five-factor model lack power to explain the long-short strategy returns. The out-of-sample portfolio cumulative log returns are plotted in Figure 1, where we find clear cross-sectional difference patterns. The return of the investment strategy mainly come from the long leg. The results show strong evidence of return predictability using BCM forecasts and the long-short investment strategy using these return forecasts generate significant profits.

[Insert Table 11 here]

[Insert Figure 1 here]

6 Conclusion

This paper presents an alternative pricing kernel to estimate and decompose the individual corporate bond risk premia. The proposed benchmark combination model (BCM) is related to the linear factor asset pricing model and the basis portfolio benchmark evaluation of Daniel et al. (1997). The BCM is a solution the high-dimensional sort difficulty in Cochrane (2011) using a linear combination of univariate-sorted basis portfolios. We provide a linear model combination framework for benchmarking individual corporate bonds, which are transparent and objective and easy to interpret the results. This approach can be easily applied to other asset classes, such as equities, commodities, and currencies.

With a non-arbitrage objective, the BCM minimizes cross-sectional pricing errors and decomposes the sources of risk premia. We apply the model to corporate bond data and find that credit ratings, short-term return reversal, and downside risk are three important sources of bond premia. In particular, short-term return reversal is an under-appreciated pricing factor that have high power for predicting bond returns.

We compare the BCM with the conventional factor models for pricing and prediction performance. The BCM has an advantage of dealing with individual assets that have a short history of data. We find evidence that the BCM model outperforms conventional factor models in pricing individual corporate bond returns. Moreover, the BCM generates return predictability for individual corporate bonds and bond basis portfolios.

References

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

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

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and E. Schaumburg (2020). Characteristic-sorted portfolios: Estimation and inference. *Review of Economics and Statistics 102*(3), 531–551.

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

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

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

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of finance 52*(3), 1035–1058.

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

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

Feng, G., N. Polson, and J. Xu (2020). Deep learning in characteristics-sorted factor models. Technical report, City University of Hong Kong.

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

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 000–000.

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

He, X., G. Feng, J. Wang, and C. Wu (2021). Predicting individual corporate bond returns. Technical report, City University of Hong Kong.

Hong, Y., H. Lin, and C. Wu (2012). Are corporate bond market returns predictable? *Journal of Banking & Finance 36*(8), 2216–2232.

Huang, T., L. Jiang, and J. Li (2020). Downside variance premium and the cross-section of corporate bond returns. Technical report, Fudan University.

Kelly, B. T., D. Palhares, and S. Pruitt (2019). Modeling corporate bond returns. Technical report, Yale University.

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

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

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

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

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

Table 2: Summary Statistics

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

	All Databases	Lehman	DataStream	TRACE&NAIC	Public	Private
Bond-month observations	589,528	126,710	63,107	399,711	124,925	464,603
Period coverage	1976-2017	1976-1998	1990-2008	1994-2017	1976-2017	1976-2017
Return - mean (%)	0.51	0.73	0.5	0.44	0.54	0.54
Return - median (%)	0.52	0.79	0.61	0.52	0.52	0.51
Excess Return - mean (%)	0.14	0.15	0.18	0.13	0.16	0.16
Excess Return - median (%)	0.22	0.27	0.25	0.18	0.14	0.13
Rating - mean	5.62	5.18	7.21	5.51	6.20	5.47
Rating - median	5	5	7	5	6	5
% of rated that are IG	89.64	93.53	84.2	89.27	88.15	90.04
% of rated that are NIG	10.36	6.47	15.8	10.73	11.85	9.96
% of Private bond	78.81	82.66	94.37	75.13	0.00	100.00
Duration - mean (years)	6.25	5.56	8.67	6.09	6.64	6.14
Duration - median (years)	5.48	5.32	9	5.06	5.82	5.39
Age - mean (years)	8.32	17.61	6.42	5.68	7.13	8.64
Age - median (year)	5.37	19.9	5.89	3.97	4.59	5.61
Amt outst. - mean ($ millions)	10,788	708	4,474	15,672	5,017	12,340
Amt outst. - median ($ millions)	20,000	1,500	2,470	30,000	2,500	15,000

Panel B: Sample Distribution By Rating & Maturity

	AAA	AA	A	BBB	Junk	All
Maturity						
2	1.87	2.66	5.45	2.78	1.11	13.87
3	1.34	2.17	4.62	2.44	1.00	11.56
4	1.34	2.06	4.61	2.39	0.98	11.37
5	0.79	1.27	3.04	1.78	0.87	7.75
6	0.80	1.16	2.92	1.77	0.86	7.51
7	0.60	1.04	2.54	1.54	0.66	6.39
8	0.59	1.03	2.47	1.56	0.59	6.24
9	0.57	1.01	2.57	1.75	0.62	6.51
10	0.16	0.53	1.07	0.88	0.33	2.96
>10	1.97	3.34	8.93	8.37	3.21	25.83
All	10.02	16.27	38.22	25.27	10.22	100.00

Panel C: Sample Distribution By Rating & Data Source

	AAA	AA	A	BBB	Junk	All
Data Sources						
Datastream	0.13	0.82	3.33	4.57	1.66	10.50
LBFI	1.44	5.99	9.11	5.01	1.43	22.99
TRACE & NAIC	8.45	9.46	25.77	15.69	7.14	66.51
All	10.02	16.27	38.22	25.27	10.22	100.00

Table 3: Predictor List

Acronym	Description	Details
		Macro Predictors
Bond market variable		
TERM	Term factor	Long-term government bond return (from Ibbotson Associates) minus the one-month Treasury bill rate
DEF	Default factor	Long-term corporate bond return minus long-term government bond return (from Ibbotson Associates)
CP5	Cochrane-Piazzesi forward factor	Codes of Monika Piazzesi, 5-year specification
ILL	Pastor-Stambaugh illiquidity	Download from Robert Stambaugh
TBL	3-month treasury bill rate	Download from Fed. St. Louis
CBMKT	Corporate bond market return	Value-weighted corporate bond market return, equal weight
MTS	Maturity spread	Return of long (greater than 10 years) maturity corporate bond returns minus return of short (2 to 5 years) maturity corporate bond returns, equal weight
RTS	Rating spread	Return of Junk bond minus return of AAA bond, equal weight
INFL	CPI index	Download from Fed. St. Louis
LTR	Long-term rate of returns	Long-term government bond return (from Ibbotson Associates)
TMS	Term spread	Long-term yield on government bonds (from Ibbotson Associates) minus the one-month Treasury bill rate
DFY	Default yield spread	Yield of BAA- corporate bond minus yield of AAA corporate bonds
Equity market variable		
DP	Dividend-to-price	S&P500 index dividend-to-price
EP	Earnings-to-price	S&P500 index earnings-to-price
NI	Net equity issuance	S&P500 index net equity issuance
LEV	Leverage	S&P500 index leverage
SVAR	Stock variance	S&P500 index variance
MKTRF	Market factor	Download from Kenneth French
SMB	Size factor	Download from Kenneth French
HML	Value factor	Download from Kenneth French
		Corporate Bond Characteristics
Fundamental		
CRT	Credit Rating	From FISD
DUR	Duration	From FISD
TMT	Time-to-maturity	From FISD
AGE	Time-from-issuance	From FISD
SIZE	Amount outstanding	From FISD
Return-distribution		
STR	Short-term Reversal	Lag 1-month return
MOM6M	6-month momentum	Lag 2-month to lag 6-month cumulative return
MOM12M	12-month momentum	Lag 2-month to lag 12-month cumulative return
LTR2Y	2-year long-term reversal	Lag 13-month to lag 24-month cumulative return
LTR3Y	3-year long-term reversal	Lag 13-month to lag 36-month cumulative return
VAR	Variance	Variance of returns of the past 36 months
DSD	Downside risk	5% VaR of returns of the past 36 months
SKEW	Skewness	Skewness of returns of the past 36 months
KURT	Kurtosis	Kurtosis of returns of the past 36 months
Covariance on risk factors		
BETA_MKT	Multiple regression beta of a five-factor model	
BETA_SMB	Multiple regression beta of a five-factor model	
BETA_HML	Multiple regression beta of a five-factor model	
BETA_DEF	Multiple regression beta of a five-factor model	
BETA_TERM	Multiple regression beta of a five-factor model	
RVAR	Residual variance in the multiple regression of a five-factor model	

Table 4: Weight (%) for Benchmark Combination

This table reports the weight (in percentage %) for the Benchmark Combination Model. Panel A is for the constrained version, and Panel B is unconstrained. The in-sample results are estimated with all observations from 1998 to 2017. We also present the subsample results for the periods of economic boom and recession. Panel A, the cells' colors are the positive frequency of weights by the 1000 bootstrap samples. Finally, we categorize the cells into four sets based on the positive frequency, divided by three breakpoints—10%, 50%, and 90%—and color the four sets of cells in the blank, light, medium, and dark backgrounds. In Panel B, the cell colors label the OLS t-stat, clustered by rating and year.

Panel A: Constrained

	In-Sample			Out-of-Sample																			
	1998-2017	Boom	Recession	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Rating	53	38	59	0	0	0	0	15	34	43	46	44	44	41	69	58	58	58	57	40	47	37	38
Duration	0	17	0	18	24	23	14	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Maturity	0	3	0	6	0	0	5	5	0	7	7	9	9	21	12	0	0	0	0	0	10	21	39
Age	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Size	0	0	0	20	22	28	33	38	25	5	0	0	0	0	0	0	0	0	0	0	0	0	0
S-term Rev	31	35	5	28	33	29	23	19	38	44	46	47	47	38	15	11	12	15	13	24	41	41	18
Mom 6M	0	0	0	15	9	5	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mom 12M	5	0	13	3	1	1	0	0	3	1	0	0	0	0	3	7	5	5	8	0	0	0	1
L-term Rev 2Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L-term Rev 3Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Variance	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Downside risk	11	0	15	0	0	0	0	0	0	0	0	0	0	0	0	21	23	21	21	36	1	0	0
Skewness	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kurtosis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beta_mkt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beta_smb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beta_hml	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beta_term	0	4	0	10	12	15	20	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beta_def	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	3	2	1	1	0	0	0	0
Residual Var	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bootstrap Positive Frequency						[0, 100)					[100, 500)					[500, 900)					[900, 1000]		

Panel B: Unconstrained

	In-Sample			Out-of-Sample																			
	1998-2017	Boom	Recession	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Rating	65	55	70	-59	-9	-3	26	43	56	60	64	65	68	72	83	67	67	67	66	50	63	69	67
Duration	0	23	-30	28	36	40	24	20	19	4	-2	-2	5	6	1	-19	-22	-27	-28	-32	5	3	18
Maturity	28	14	21	16	7	5	11	16	3	33	40	45	42	52	45	26	24	26	26	32	44	51	36
Age	-42	-31	-17	-17	-15	-20	-25	-31	-43	-61	-52	-54	-54	-48	-18	-28	-27	-25	-25	-40	-36	-31	-15
Size	25	14	24	35	35	41	42	47	45	34	26	23	11	8	26	18	19	21	20	13	33	29	27
S-term Rev	48	50	26	42	44	42	35	35	53	58	62	65	69	67	43	30	32	35	34	42	63	63	49
Mom 6M	7	10	7	26	21	20	19	-8	-5	-2	-4	-5	3	0	8	10	10	11	11	14	11	16	11
Mom 12M	19	16	20	23	20	23	18	15	31	29	29	31	28	13	22	15	13	13	16	13	9	11	18
L-term Rev 2Y	-7	2	-20	-1	-6	8	3	6	-2	-7	-7	4	-1	8	1	-10	-11	-10	-12	-9	10	6	2
L-term Rev 3Y	-19	-6	-38	2	-2	1	2	-12	-17	-16	-19	-21	-13	-8	-34	-24	-19	-23	-21	-6	-6	-21	-9
Variance	4	9	-13	-44	-44	-45	-23	-15	0	9	13	14	18	18	-2	-2	-5	-6	-9	-1	0	8	-9
Downside risk	18	1	38	10	6	-4	-9	-9	-2	-3	-3	-5	-9	-11	4	34	36	35	36	35	16	5	15
Skewness	5	-9	17	8	12	31	30	39	30	17	5	-4	-27	-45	-5	10	10	10	11	13	-18	-23	-33
Kurtosis	-26	-28	-3	-10	-25	-40	-38	-21	-25	-20	-15	-18	-22	-19	7	-17	-15	-16	-15	-41	-50	-45	-27
Beta_mkt	-15	-16	-10	6	3	-7	2	-6	-19	-26	-22	-26	-23	-18	-16	-11	-11	-9	-8	-4	-9	-4	-2
Beta_smb	-19	-12	-13	14	7	-6	-26	-26	-3	-10	-8	-4	-3	-14	-11	-17	-16	-15	-15	-20	-19	-22	-42
Beta_hml	-2	-4	5	10	3	6	-6	2	14	16	12	12	16	4	-54	3	2	2	1	18	-26	-26	-30
Beta_term	6	21	4	18	25	33	37	32	12	13	14	14	15	17	2	3	0	-4	-5	3	-1	9	17
Beta_def	5	-19	31	-5	-13	-17	-24	-29	-51	-44	-45	-46	-34	-16	16	26	25	26	26	25	3	3	-5
Residual Var	9	18	-7	5	2	3	12	12	16	24	23	23	19	22	-8	-4	0	2	3	10	18	9	18
Absolute t-stat						[0, 1.645)					[1.645, 1.960)					[1.960, 2.576)					[2.576,)		

Table 5: Pricing Performance for Individual Corporate Bonds

This table reports the pricing performance for individual corporate bond returns. BCMc is the benchmark combination model with constraints, and BCMu is the one without constraints. The FF5 denotes the five-factor model in Fama and French (1993) and the five factors are MKT, SMB, HML, TERM, and DEF. The BBW4 denotes the factor model proposed in Bai et al. (2019). We compare the performance between BCM and factor models. For example, BCMc-FF5 indicates a performance improvement of BCMc against FF5. Both Panels A and B report the performance for the total R^2 in Eq. (11). Panel C reports the predictive R^2 in Eq. (12). Each of the R^2 columns reports the aggregate pricing performance, pooling all bonds and periods. FM-$\overline{R^2}$ reports the time-series average of $\{R_t^2\}_{t=1}^T$. We also report the Fama-MacBeth t-test for FM-$\overline{R^2}$, where the signs ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	IG		NIG		Overall	
	R^2	FM-$\overline{R^2}$	R^2	FM-$\overline{R^2}$	R^2	FM-$\overline{R^2}$
Panel A: Total R^2 %, In-Sample						
BCMc	17.96	16.35***	13.20	11.77***	16.00	15.03***
BCMu	19.25	18.05***	14.07	12.59***	17.12	16.52***
FF5	10.02	-6.30**	17.88	-8.52	13.26	-5.05
BBW4	18.49	8.26***	22.24	7.25***	20.03	9.01***
BCMc-FF5	7.94	22.65***	-4.68	20.29***	2.74	20.08***
BCMc-BBW	-0.53	8.09***	-9.04	4.52***	-4.03	6.02***
BCMu-FF5	9.23	24.36***	-3.81	21.11***	3.86	21.57***
BCMu-BBW	0.76	9.79***	-8.17	5.35***	-2.91	7.52***
Panel B: Total R^2 %, Out-of-Sample						
BCMc	17.91	16.77***	12.24	11.34***	15.57	15.17***
BCMu	17.65	17.47***	12.87	11.99***	15.68	15.98***
FF5	-12.60	-10.66***	-14.01	-28.35***	-13.18	-14.26***
BBW4	-6.00	0.45	-8.51	-8.22***	-7.04	-0.80
BCMc-FF5	30.51	27.43***	26.25	39.69***	28.75	29.43***
BCMc-BBW	23.91	16.32***	20.75	19.56***	22.61	15.97***
BCMu-FF5	30.25	28.13***	26.88	40.34***	28.86	30.24***
BCMu-BBW	23.65	17.02***	21.38	20.22***	22.72	16.78***
Panel C: Predictive R^2 %						
BCMc	0.49	1.26***	0.41	1.22***	0.45	1.24***
BCMu	0.72	1.69***	0.51	1.56***	0.63	1.63***
FF5	-0.22	0.35	0.16	0.26	-0.06	0.39
BBW4	-0.43	0.12	0.19	0.21	-0.17	0.22
BCMc-FF5	0.71	0.91***	0.25	0.97***	0.51	0.85***
BCMc-BBW	0.92	1.14***	0.22	1.01***	0.62	1.02***
BCMu-FF5	0.94	1.35***	0.35	1.31***	0.69	1.23***
BCMu-BBW	1.15	1.57***	0.32	1.35***	0.8	1.41***

Table 6: Fama-MacBeth Predictive $\overline{R^2}$ % by 5×5 Rating-Duration Style Box

This table reports the Fama-MacBeth predictive $\overline{R^2}$ of the 5×5 rating-duration sorts. Panel A is for the benchmark combination model (constrained). Panel B uses the average returns of the corresponding bivariate-sorted portfolios as a benchmark to individual corporate bond returns. Panel C reports the Fama-MacBeth t-statistics for the differences between these two models. The signs ***, **, and * indicate t-stat significance at the 1%, 5%, and 10% level, respectively.

Rating	AAA	AA	A	BBB	NIG
			Panel A: Benchmark Combination Model (Constrained)		
Duration					
1	-0.54	-1.55	0.08	0.31	0.94***
2	0.96	1.66***	1.79***	1.34***	1.46***
3	1.35***	1.31***	1.47***	1.37***	1.18***
4	1.92***	1.87***	1.77***	1.38***	1.47***
5	2.13***	1.49***	1.30***	1.69***	1.12***
			Panel B: Bivariate-Sorted Benchmark on Rating and Duration		
Duration					
1	0.22	-0.11	-0.74	-0.84	0.35
2	0.76	0.85***	1.03***	0.54*	0.88**
3	0.54**	0.48**	0.75***	0.68***	0.86***
4	0.54*	0.85***	1.00***	0.67***	1.17***
5	1.25**	0.80*	0.18	1.36***	0.44**
			Panel C: Difference A Minus B		
Duration					
1	-0.76	-1.44	0.82***	1.15***	0.59***
2	0.19	0.81***	0.76***	0.79***	0.59***
3	0.81***	0.83***	0.72***	0.69***	0.32***
4	1.38***	1.03***	0.77***	0.71***	0.30***
5	0.88***	0.69***	1.12***	0.33*	0.68**

Table 7: Pricing Performance for Private and Public Individual Bonds

This table reports the BCM pricing performance for pricing private bonds and public bonds. Each of the R^2 columns reports the aggregate pricing performance, pooling all bonds and periods. FM-$\overline{R^2}$ reports the time-series average of $\{R_t^2\}_{t=1}^T$, and the last column is the difference between private and public bonds. The table format and performance measures follow the main results in Table 5.

		Private		Public		Difference
		R^2	FM-$\overline{R^2}$	R^2	FM-$\overline{R^2}$	FM-$\overline{R^2}$
Panel A: Total R^2 %, In-Sample						
BCMc	IG	16.85	15.36***	22.37	20.19***	-4.83***
	NIG	12.60	11.68***	14.91	12.40***	-0.72
	ALL	15.17	14.22***	18.84	18.10***	-3.88***
BCMu	IG	18.10	16.91***	23.83	22.37***	-5.46***
	NIG	13.44	12.51***	15.89	12.90***	-0.39
	ALL	16.26	15.60***	20.07	19.93***	-4.33***
Panel B: Total R^2 %, Out-of-Sample						
BCMc	IG	16.84	15.79***	22.15	20.54***	-4.75***
	NIG	11.78	11.37***	13.56	11.91***	-0.55
	ALL	14.85	14.42***	18.09	18.09***	-3.69***
BCMu	IG	16.66	16.37***	21.59	21.54***	-5.17***
	NIG	12.37	11.83***	14.31	12.50***	-0.67
	ALL	14.97	15.09***	18.15	19.24***	-4.84***
Panel C: Predictive R^2 %						
BCMc	IG	0.47	1.21***	0.54	1.53***	-0.32**
	NIG	0.41	1.14***	0.43	1.70***	-0.56***
	ALL	0.45	1.17***	0.49	1.56***	-0.39***
BCMu	IG	0.72	1.65***	0.72	1.95***	-0.29
	NIG	0.49	1.47***	0.55	2.15***	-0.68**
	ALL	0.63	1.56***	0.64	1.98***	-0.42**
Panel E: Total R^2 %, In-Sample						
BCMc	1998-2007	13.16	12.37***	13.22	13.08***	-0.70
	2008-2009	17.58	14.93***	23.72	14.93***	-4.66***
	2010-2017	16.90	16.35***	27.86	24.01***	-7.66***
BCMu	1998-2007	14.34	13.66***	14.27	14.19***	-0.53
	2008-2009	18.37	15.79***	24.44	15.79***	-4.87***
	2010-2017	18.34	17.97***	31.17	26.93***	-8.96***
Panel F: Total R^2 %, Out-of-Sample						
BCMc	1998-2007	13.18	12.89***	12.35	13.01***	-0.12
	2008-2009	16.66	14.30***	22.81	18.84***	-4.53***
	2010-2017	16.69	16.35***	28.00	24.25***	-7.90***
BCMu	1998-2007	13.92	13.57***	13.13	13.88***	-0.32
	2008-2009	15.46	13.59***	20.95	17.86***	-4.27**
	2010-2017	17.67	17.38***	30.31	26.28***	-8.90***
Panel G: Predictive R^2 %						
BCMc	1998-2007	0.24	0.26**	0.13	0.30**	-0.03
	2008-2009	0.17	0.24*	0.21	0.44**	-0.20**
	2010-2017	1.84	2.54***	2.58	3.42***	-0.88***
BCMu	1998-2007	0.42	0.45***	0.27	0.45***	-0.01
	2008-2009	0.21	0.30*	0.24	0.56**	-0.25*
	2010-2017	2.40	3.27***	3.17	4.23***	-0.97**

Table 8: Predicting Corporate Bond Basis Portfolio Returns

This table reports the out-of-sample R^2_{OOS} (%) of return forecasts for basis bond portfolios. The prediction methods are listed for each column, and the out-of-sample prediction baseline is the average return of each basis portfolio. We aggregate the results for the five basis portfolios of all 20 characteristics.

Characteristic	MeanComb	Lasso	PCA	Random Forest
Rating	2.17	9.19	13.52	3.34
Duration	3.35	16.25	17.43	8.82
Maturity	3.42	17.62	18.30	8.37
Age	3.60	17.76	18.03	11.09
Size	3.61	15.09	15.75	6.87
Short-term Rev	3.50	17.66	16.32	14.05
Momentum 6M	2.85	14.64	14.65	6.84
Momentum 12M	2.76	16.51	12.07	6.72
Long-term Rev 2Y	3.88	16.39	22.03	14.17
Long-term Rev 3Y	4.12	18.86	20.67	13.74
Variance	3.55	10.96	19.52	13.22
Downside risk	3.14	11.06	16.68	10.59
Skewness	3.98	18.97	16.64	12.80
Kurtosis	4.09	17.08	19.13	13.10
Beta_mkt	3.28	16.59	17.35	9.89
Beta_smb	3.67	13.06	16.80	10.86
Beta_hml	3.44	12.69	16.45	6.63
Beta_term	3.66	17.74	12.92	9.78
Beta_def	3.31	16.35	14.39	11.92
Residual Var.	3.56	16.41	19.16	12.53

Table 9: Predicting Individual Bond Returns with Basis Portfolio Return Forecasts

This table reports the out-of-sample R^2_{OOS} (%) of predicting individual corporate bond returns with the basis portfolio return forecasts. We identify each characteristic's quintile portfolio for each bond and use the basis portfolio return forecasts to predict the bond's returns. For example, in the first row, we find the rating basis portfolio to which the bond belongs and use the tating basis portfolio return forecast to predict the bond return. In each row, we use a group of five basis portfolios on a single bond characteristic. In the columns, we have four methods to predict portfolio returns: (1) Mean Combination, (2) Lasso, (3) PCA regression, and (4) Random Forest. The out-of-sample R^2_{OOS} baseline calculation is the belonging rating basis portfolio average return. For comparison, we also report the BCM forecasts in the top panel. The BCM forecasts combine the 20 basis portfolio forecasts, and more details are reported in Table 10. We report the Fama-MacBeth t-statistics, where the signs ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Average		MEAN Combination		Lasso		PCA Regression		Random Forest	
Characteristic	R^2	FM-R^2	R^2	FM-R^2	R^2	FM-R^2	R^2	FM-R^2	R^2	FM-R^2
BCMc	0.27	0.64***	0.62	1.02***	2.43	1.75***	2.19	1.12	1.68	2.18***
BCMu	0.46	1.06***	0.82	1.54***	2.42	1.94***	2.04	0.97	1.58	2.27***
Rating	0.00	0.00	0.36	0.02	1.41	0.31	1.83	0.58	0.79	0.95*
Duration	0.01	0.13*	0.36	0.13*	1.77	0.54	1.67	0.34	1.06	1.18***
Maturity	0.02	0.12**	0.37	0.15**	1.92	0.87	1.76	0.34	1.06	1.13**
Age	-0.02	-0.03	0.32	-0.02	1.82	0.69	1.61	0.14	1.20	1.20***
Size	-0.01	0.05	0.37	0.10	1.78	0.64	1.55	0.09	0.98	1.16***
Short-term Rev	0.60	1.09***	0.98	1.06***	2.65	1.96***	2.19	1.04	2.36	2.81***
Momentum 6M	-0.02	0.00	0.28	-0.04	1.68	0.85	1.42	0.23	0.78	1.31***
Momentum 12M	-0.03	-0.06	0.28	-0.07	1.96	0.97*	1.21	0.15	0.88	1.35***
Long-term Rev 2Y	0.00	-0.21	0.40	-0.20	1.80	0.47	2.23	0.59	1.53	1.18***
Long-term Rev 3Y	-0.02	-0.14	0.41	-0.13	1.99	0.73	2.08	0.65	1.47	1.13**
Variance	0.06	0.21***	0.46	0.21***	1.42	-0.52	2.22	-0.12	1.81	1.35***
Downside risk	0.14	0.26***	0.50	0.26***	1.72	0.50	1.97	-0.34	1.54	1.38***
Skewness	0.10	-0.14	0.49	-0.14	2.05	0.47	1.66	-0.63	1.51	1.40***
Kurtosis	-0.01	-0.09	0.37	-0.08	1.73	0.22	1.72	0.24	1.40	1.29***
Beta_mkt	0.00	0.03	0.34	0.02	1.78	0.64	1.71	0.06	1.08	1.04**
Beta_smb	0.00	0.02	0.36	0.03	1.38	-0.40	1.57	-0.13	1.16	1.30***
Beta_hml	0.01	0.03	0.35	0.03	1.36	0.30	1.58	0.30	0.87	0.78*
Beta_def	-0.01	-0.03	0.38	0.00	1.95	0.48	1.30	-0.49	1.21	0.94**
Beta_term	0.01	0.04	0.35	0.02	1.75	0.45	1.43	-0.41	1.33	0.94**
Residual Variance	0.05	0.15***	0.43	0.15**	1.99	0.83	2.10	-0.11	1.78	1.46***

Table 10: Predicting Individual Bond Returns with the BCM Forecast

This table reports out-of-sample R^2_{OOS} (%) of predicting individual corporate bond returns using the BCM forecast. We present results to both investment-grand and non-investment-grade bonds and both pre- and post-crisis periods. In Panel A, B, D, and E, the predictions are performed by the BCM forecasts, where the common combination weights are reported in Table 4. Panel C and F report the prediction performance of multifactor models. The table format and performance measures follow the main results in Table 9.

	IG		NIG		ALL	
	R^2	FM-$\overline{R^2}$	R^2	FM-$\overline{R^2}$	R^2	FM-$\overline{R^2}$
Panel A: Benchmark Combination Model, Constrained						
c-Avg.	0.39	0.81***	0.09	0.41***	0.27	0.64***
c-MeanComb	0.95	1.25***	0.15	0.69***	0.62	1.02***
c-Lasso	3.08	2.22***	1.50	0.47	2.43	1.75***
c-PCA	2.56	1.19	1.66	0.36	2.19	1.12
c-RandomForest	2.66	2.83***	0.27	0.44	1.68	2.18***
Panel B: Benchmark Combination Model, Unconstrained						
u-Avg.	0.65	1.30***	0.19	0.77***	0.46	1.06***
u-MeanComb	1.27	1.85***	0.17	1.11***	0.82	1.54***
u-Lasso	3.22	2.58***	1.27	0.17	2.42	1.94***
u-PCA	2.49	1.06	1.41	-0.06	2.04	0.97
u-RandomForest	2.77	3.05***	-0.12	0.10	1.58	2.27***
Panel C: Factor Model						
FF5	-0.46	-0.34*	-0.08	-0.46*	-0.31	-0.35*
BBW4	-0.68	-0.54***	-0.05	-0.50**	-0.42	-0.50***

	1998-2007		2008-2009		2010-2017	
	R^2	FM-$\overline{R^2}$	R^2	FM-$\overline{R^2}$	R^2	FM-$\overline{R^2}$
Panel D: Benchmark Combination Model, Constrained						
c-avg.	0.30	0.49***	0.09	0.21***	0.59	0.93***
c-MeanComb	0.47	0.83***	0.47	0.56***	1.57	1.38***
c-Lasso	1.82	2.44***	3.31	3.30***	2.60	0.49
c-PCA	0.76	1.77***	4.34	4.10***	2.30	-0.43
c-RandomForest	2.00	2.56***	0.53	1.44***	3.28	1.90**
Panel E: Benchmark Combination Model, Unconstrained						
u-avg.	0.48	0.67***	0.14	0.29***	1.17	1.74***
u-MeanComb	0.66	1.03***	0.44	0.58***	2.32	2.42***
u-Lasso	2.05	2.59***	2.95	2.84**	2.49	0.89
u-pca	0.66	1.69**	4.25	3.96***	1.85	-0.68
u-RandomForest	2.02	2.49***	0.22	1.18**	3.28	2.27**
Panel F: Factor Model						
FF5	-0.57	-0.75***	-0.08	-0.53	0.12	0.20
BBW4	-0.74	-0.97***	-0.09	-0.56	-0.03	0.10

Table 11: Forecast-Implied Long-Short Strategy Performance

This table reports the performance for return forecast-implied long-short strategies. The performance measures include average excess returns (%), alphas (%) on a five-factor model (MKT, SMB, HML, TERM, DEF), t-stat for the alpha, and the annualized Sharpe ratios. The signs ***, **, and * indicate the significance of alphas at the 1%, 5%, and 10% level, respectively.

	IG			NIG			ALL		
Method	Avg.Ret	FF5-α	SR	Avg.Ret.	FF5-α	SR	Avg.Ret.	FF5-α	SR
cbmkt	0.23	0.07	0.66	0.49	0.26*	0.61	0.27	0.10*	0.73
Panel A: Benchmark Combination Model, Constrained									
c-avg.	0.99	0.99***	3.22	1.19	1.07***	1.41	0.94	0.93***	2.57
c-MeanComb	1.05	1.04***	3.29	1.07	0.91***	1.34	0.91	0.89***	2.20
c-Lasso	0.93	0.92***	2.73	1.14	1.02***	1.35	0.82	0.81***	1.81
c-PCA	0.76	0.74***	2.25	0.83	0.72***	1.01	0.73	0.71***	1.51
c-RandomForest	0.92	0.87***	2.84	1.12	0.99***	1.55	0.78	0.73***	1.73
Panel B: Benchmark Combination Model, Unconstrained									
u-avg.	0.85	0.75***	2.24	1.30	1.15***	1.36	0.93	0.82***	1.93
u-MeanComb	0.89	0.74***	2.32	1.08	0.84***	1.19	0.86	0.69***	1.74
u-Lasso	0.75	0.67***	2.43	1.17	1.09***	1.64	0.73	0.66***	1.70
u-pca	0.59	0.48***	1.63	0.46	0.30*	0.57	0.62	0.51***	1.25
u-RandomForest	0.71	0.58***	2.05	0.75	0.56***	0.82	0.70	0.55***	1.44

Figure 1: Cumulative Log Return of Benchmark Combination Forecast & Machine Learning

This figure shows the cumulative log return of the return forecast-implied portfolios. The performance for these return forecasts is reported in Table 10, where 5 indicates the forecast winner and 1 the loser. We present the results for the weighted BCM forecast of four predictive models (Mean Combination, Lasso, PCA regression, and Random Forest).