Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Forecasting Corporate Bond Returns with a Large Set of Predictors: An Iterated Combination Approach

Hai Lin, Chunchi Wu, Guofu Zhou

To cite this article:
Hai Lin, Chunchi Wu, Guofu Zhou (2017) Forecasting Corporate Bond Returns with a Large Set of Predictors: An Iterated Combination Approach. Management Science

Published online in Articles in Advance 10 May 2017

. http://dx.doi.org/10.1287/mnsc.2017.2734 http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–21
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Forecasting Corporate Bond Returns with a Large Set of Predictors: An Iterated Combination Approach

Hai Lin,[a] **Chunchi Wu,**[b] **Guofu Zhou**[c]

[a] Victoria University of Wellington, Kelburn, Wellington 6012, New Zealand; [b] State University of New York at Buffalo, Buffalo, New York 14228; [c] Olin School of Business, Washington University in St. Louis, St. Louis, Missouri 63130

Contact: hai.lin@vuw.ac.nz (HL); chunchiw@buffalo.edu (CW); zhou@wustl.edu (GZ)

Received: November 20, 2015
Accepted: October 29, 2016
Published Online in Articles in Advance: May 10, 2017

https://doi.org/10.1287/mnsc.2017.2734

Abstract. Using a comprehensive return data set and an array of 27 macroeconomic, stock, and bond predictors, we find that corporate bond returns are highly predictable based on an iterated combination model. The large set of predictors outperforms traditional predictors substantially, and predictability generated by the iterated combination is both statistically and economically significant. Stock market and macroeconomic variables play an important role in forming expected bond returns. Return forecasts are closely linked to the evolution of real economy. Corporate bond premia have strong predictive power for business cycle, and the primary source of this predictive power is from the low-grade bond premium.

History: Accepted by Lauren Cohen, finance.
Funding: H. Lin acknowledges the support of the National Natural Science Fund of China [Grants 71301137, 71371161, and 71101121].
Supplemental Material: The Internet appendix is available at https://doi.org/10.1287/mnsc.2017.2734.

Keywords: predictability • corporate bonds • iterated combination • out-of-sample forecasts • utility gains

There is a large body of literature on whether stock returns are predictable, and there is also an equally impressive number of studies on government bond returns, but there is only a handful of research on the predictability of corporate bond returns.[1] The vast studies have considerably improved our understanding for time variations in equity and government bond risk premia and their roles in asset pricing, portfolio allocation, risk management, and performance evaluation of investment managers. By contrast, much less is known about corporate bond risk premia. Keim and Stambaugh (1986) conduct perhaps the first major study on predicting corporate bond returns. Subsequently, Fama and French (1989) find that default spreads, term spreads, and dividend yields are valuable predictors both in-sample and out-of-sample. More recently, Greenwood and Hanson (2013) and Lin et al. (2014) further identify issuer quality and liquidity and forward rate factors, separately, as additional predictors for corporate bond returns. Variant to these time-series predictability studies, Chordia et al. (2016) and Choi and Kim (2016) study the cross-sectional predictability of corporate bond returns, investigating whether equity variables that capture stock return anomalies can explain the cross section of expected bond returns. The momentum studies of Jostova et al. (2013) and Lin et al. (2016) are also on cross-sectional predictability.

In this paper, we conduct a comprehensive study on the time-series predictability of corporate bond returns using both a large data set and a new forecast method. This line of research is important because it helps understand the time-varying risk premia in the corporate bond market, which is a vital sector of the financial system with a sheer size of about 10 trillion dollars and is the primary source of long-term capital in the United States (Bhojraj and Sengupta 2003). From the perspective of investors and fund managers, corporate bond return predictability is of fundamental importance for asset pricing and portfolio allocations. Moreover, studying bond risk premia is essential for understanding firms' interest rate exposure as well as corporate financing choices and capital structure. Finally, to the extent that time-varying bond risk premia carry information for forecasting the risk-bearing capacity of the financial sector, corporate bond return forecasts provide important signals for future aggregate financial risk.

We address four major questions. The first is what economic variables can have predictive power for corporate bond returns. The number of predictors used by Fama and French (1989) and Greenwood and Hanson (2013), among others, is admittedly few. While these studies provide insights into why certain predictors should be looked at, they ignore other potentially important predictors and hence can underestimate the true predictability of bond returns. Indeed, the economic value of predictability from a limited number of predictors prescribed by the existing studies is not significant. From an investment perspective, it is important to find the maximum predictability of returns because underestimated predictability can lead to less investment gains. Toward this goal, it is necessary to explore a number of predictors that have potentially important information for return forecasts. In this paper, we consider three types of predictors that are relevant in theory for corporate bond returns: stock market, Treasury market, and corporate bond market variables. The predictors are those typical in the literature. For example, Chordia et al. (2016) and our paper share most of the stock predictors, though we use the aggregate ones. A unique approach taken here is that we study the joint predictability of all the predictors, a total of 27, simultaneously. As we shall demonstrate, the pooling of information from the large set of predictors improves the predictability of corporate bond returns dramatically.[2]

The second question is how to combine the information from a large set of predictors to obtain optimal bond return forecasts. As shown by Welch and Goyal (2008) in the context of equity risk premium forecasts, a naive multiple regression of asset returns on a large number of predictors will overparameterize the model and lead to poor out-of-sample forecasts. While the principal component analysis (PCA) is a popular method in the literature for extracting information from a large number of variables, it does not perform well out of sample either in our applications. A well-known econometric tool (see Timmermann 2006) is a combination method. The predictive regressions are first run on each predictor to obtain individual forecasts, and then a combination of the individual forecasts, such as their mean, serves as the forecast. In macroeconomic forecasting, Stock and Watson (2001) find that such a simple mean combination (MC) method is the favored strategy rather than using dozens of individual predictive models. Consistent with their finding, Rapach et al. (2010) show that the MC delivers a significant out-of-sample forecast of the equity risk premium. In this paper, we use the MC as well as a weighted combination (WC) proposed by Bates and Granger (1969).

A unique methodological contribution of this paper is that we provide a simple method to improve the MC and WC further by combining them again with the historical sample mean forecast. From an econometric standpoint, the new combined forecasts are a special case of the general combination framework set out by Granger and Ramanathan (1984) and Capistrán and Timmermann (2009), adapted to our applications. Intuitively, as the second-step combination combines either MC or WC optimally with the historical sample mean, it should generally provide better forecasts than using either MC (or WC) or the historical mean alone. To emphasize this feature of repeated combination, we call the new combinations the iterated mean combination (IMC) and the iterated weighted combination (IWC), respectively.

We show that the IMC forecast has a close relationship to the partial least squares (PLS) forecast, which was first proposed by Wold (1966) and has recently been developed further by Kelly and Pruitt (2013, 2015).[3] The PLS and the IMC, in fact, belong to the same class of forecasts in the case of linear models, though the latter is more general and applicable to nonlinear models as well. Hence, our proposed methodology not only advances the literature on combination forecasts but also provides an alternative interpretation for the powerful PLS forecast. Since our applications show that the IWC can improve the IMC further, the IWC forecast will be our focus throughout this paper.

The third question is whether the predictability is of economic value. While Fama and French (1989) and Greenwood and Hanson (2013) find that the predictability of corporate bond returns is statistically significant, the issue of economic significance is not investigated. Considering an investor who has a mean-variance utility with a risk aversion of 5, we find that the average utility gains (annualized certainty equivalent returns) from ignoring the predictability completely to using the predictability based on the IWC are 5.74% (3.77%) at the monthly (quarterly) horizon. By contrast, the average gains are less than 1.86% for the best existing model, the Fama–French (1989) model, at both monthly and quarterly horizons, which are not economically significant at the conventional 2% cutoff point. Thus, our use of comprehensive bond and stock predictors and the new methodology produces distinctly better out-of-sample forecasts, which are not only statistically but economically significant. Furthermore, the economic gain of using the proposed prediction methodology is robust to transaction costs.

The fourth question is what the economic sources are that drive the corporate bond return predictability. Fama and French (1989) are the first to link variations in expected corporate bond returns to business conditions. However, their inference is based only on in-sample forecasts, and it is unclear whether or not the out-of-sample forecasts are also tied to business conditions. We conduct extensive analysis on the economic sources of out-of-sample corporate bond return predictability. Our results, confirming Fama and French's (1989) in-sample study that return predictability of corporate bonds is linked to variations in business conditions, suggest that time-varying macroeconomic risk is the main source of return predictability.

Our paper is about time-series predictability that explains the time-varying bond risk premia. This is different from cross-sectional predictability that is about predictability of one group of bonds relative to another in the cross section. For example, Chordia et al. (2016)

examine whether equity variables, such as firm size, profitability, and idiosyncratic volatility, can explain bond return differences, and Choi and Kim (2016) find that asset growth and investment can predict bond returns cross-sectionally. By contrast, our paper focuses on the traditional time-series predictability (e.g., Fama and French 1989)—that is, the predictability of bond risk premia over time—and the predictors in our study are all aggregate variables rather than at the firm level. Moreover, we examine time variations in the aggregate bond risk premia and focus on out-of-sample forecasts and the economic gains from the time-series forecasts. In addition, we link the bond risk premia to changing macroeconomic risk. In short, as in equity studies, time-series predictability and cross-sectional predictability are quite different, but both provide valuable insights into understanding the behavior of expected bond returns.

1. The Methodology

In this section, we introduce a new econometric methodology that pools the information from a large set of predictors and has a linkage with the PLS forecast recently advanced by Kelly and Pruitt (2013).

1.1. Standard Combinations

In forecasting future corporate bond excess returns, we use the standard predictive regression model:

$$r_{t+1} = \alpha + \beta_1 z_{1t} + \beta_2 z_{2t} + \cdots + \beta_N z_{Nt} + \varepsilon_{t+1}, \tag{1}$$

where r_{t+1} is the return of a corporate bond in excess of the riskless rate, z_{jt} is the jth predictor at time t ($j = 1, \ldots, N$), and ε_{t+1} is an error term with mean equal to zero. For the Fama–French (1989, hereafter referred to as FF) model, $N = 2$ if the predictors are term spreads and default spreads (or $N = 3$ if dividend yields are also included).

When N is large, the predictive regression model is generally poorly behaved because of limited data in practice. For example, when $N = 14$, Welch and Goyal (2008) show that the "kitchen sink" regression with all predictors ends up with useless out-of-sample forecasts for the equity risk premium. In our case with $N = 27$, the problem is further compounded.

A practical solution is to use forecast combination methods (e.g., Timmermann 2006). The idea is first to run the predictive regression on each predictor

$$r_{t+1} = a_j + b_j z_{jt} + \varepsilon_{t+1,\,j} \tag{2}$$

to obtain individual forecasts,

$$\hat{r}_{t+1\,|\,t,\,j} = \hat{a}_j + \hat{b}_j z_{jt}, \tag{3}$$

where $\hat{a}_j$ and $\hat{b}_j$ are the regression coefficients from the individual predictive regression on the jth predictor, where $\varepsilon_{t+1,\,j}$ is, as usual, the disturbance with a mean equal to zero, and then combine the individual forecasts. A mean combination forecast will be the average of the N individual forecasts that utilizes the information of all predictors,

$$\hat{r}^{MC}_{t+1\,|\,t} = \frac{1}{N}\hat{r}_{t+1\,|\,t,1} + \frac{1}{N}\hat{r}_{t+1\,|\,t,2} + \cdots + \frac{1}{N}\hat{r}_{t+1\,|\,t,N}. \tag{4}$$

Besides the mean combination or average forecast, the median and trimmed mean combinations are also often used. Bates and Granger (1969) propose another simple combination method that sets the combination weights to be proportional to the inverse of estimated residual variances, known as the weighted-average forecast,

$$\hat{r}^{WC}_{t+1\,|\,t} = \frac{1/\hat{\sigma}^2_{t,1}}{\sum_{j=1}^{N}(1/\hat{\sigma}^2_{t,j})}\hat{r}_{t+1\,|\,t,1} + \frac{1/\hat{\sigma}^2_{t,2}}{\sum_{j=1}^{N}(1/\hat{\sigma}^2_{t,j})}\hat{r}_{t+1\,|\,t,2} + \cdots + \frac{1/\hat{\sigma}^2_{t,N}}{\sum_{j=1}^{N}(1/\hat{\sigma}^2_{t,j})}\hat{r}_{t+1\,|\,t,N}, \tag{5}$$

where $\hat{\sigma}^2_{t,j}$s are the estimated residual variance from the individual predictive regressions (2) using information up to time t. If individual forecasts are unbiased, both mean combination and weighted-average combination forecasts are unbiased. Both $\hat{r}^{MC}_{t+1\,|\,t}$ and $\hat{r}^{WC}_{t+1\,|\,t}$ will be used in this paper.

Both $\hat{r}^{MC}_{t+1\,|\,t}$ and $\hat{r}^{WC}_{t+1\,|\,t}$ are based on the simple weighting of individual forecasts. Although there are many alternatives in the literature (see Timmermann 2006 and references therein), later studies (see, e.g., Rapach et al. 2010 and references therein) show that the simple combination forecasts work well in practice and often do better than complex ones, especially when the number of predictor N is large or the time-series data points T are relatively small. Hence, our paper will focus on $\hat{r}^{MC}_{t+1\,|\,t}$ and $\hat{r}^{WC}_{t+1\,|\,t}$, which are simple and generally reliable. We also consider median ($\hat{r}^{MD}_{t+1\,|\,t}$) and trimmed mean ($\hat{r}^{TC}_{t+1\,|\,t}$) combination as a robustness check. The median combination forecast selects the median of forecasts by N predictors, and the trimmed mean combination forecast is the mean forecast by excluding the largest and smallest values of individual forecasts.

1.2. Iterated Combinations

Unlike existing studies, our paper proposes an iterated combination—that is, a further combination of the combination forecasts of either $\hat{r}^{MC}_{t+1\,|\,t}$ or $\hat{r}^{WC}_{t+1\,|\,t}$ with $\bar{r}_t$, which is the sample mean of r_{t+1} using all observations till time t. Consider, for example, the combination of $\hat{r}^{MC}_{t+1\,|\,t}$ with $\bar{r}_t$. From a statistical standpoint, we are interested in a predictor of the following regression type:

$$r_{t+1} = b_0 + b_1\bar{r}_t + b_2\hat{r}^{MC}_{t+1\,|\,t} + u_{t+1}, \tag{6}$$

where u_{t+1} is the noise. Instead of using the bivariate regression, we use the constrained version,

$$r_{t+1} = (1-\delta)\bar{r}_t + \delta\hat{r}^{MC}_{t+1\,|\,t} + u_{t+1}, \tag{7}$$

because it has an interesting portfolio diversification interpretation.[4] A suitable portfolio of $\bar{r}_t$ and $\hat{r}^{MC}_{t+1|t}$ is generally better than using either $\bar{r}_t$, the popular benchmark, or $\hat{r}^{MC}_{t+1|t}$, the conventional one-step combination forecast. Since our combination method essentially repeats a combination to a combination forecast, we refer to it as an *iterated* combination method.

Granger and Ramanathan (1984) provide three versions of regressions to combine forecasts that include the above regression as a special case. Capistrán and Timmermann (2009) extend their framework further with a time-varying number of predictors and provide simulation evidence. The study by Capistrán and Timmermann (2009) seems to be the first and the only other study that considers an iterated combination that also combines the mean and the average forecast. However, in their applications, Capistrán and Timmermann (2009) are more concerned about the bias of the average forecasts of experts and so they use the mean to correct this bias. By contrast, the average forecast here, obtained from averaging the regression forecasts with an intercept, is unbiased already. Hence, bias reduction is not an issue here. Instead, our primary objective is to increase the R^2 or to minimize the mean squared error.

Mathematically, our objective is to solve the following optimization problem:

$$\min_{\delta} E_t(r_{t+1} - \hat{r}_{t+1|t})^2 = E_t\left[r_{t+1} - (1-\delta)\bar{r}_t - \delta \hat{r}^{MC}_{t+1|t}\right]^2. \quad (8)$$

Note that although in principle δ can be any real number, it is generally greater than 1 in our applications below. In the special cases, $\delta = 0$ implies that $\hat{r}^{MC}_{t+1|t}$ has no information whatsoever, and $\delta = 1$ suggests that it is unnecessary to use information about $\bar{r}_t$ to improve $r^{MC}_{t+1|t}$. Theoretically, there exists such a δ that makes the new combination better than either $\bar{r}_t$ or $r^{MC}_{t+1|t}$. Indeed, it is easy to solve the optimal δ from the first-order condition of the objective function:

$$\delta^* = \frac{\mathrm{cov}_t(r_{t+1} - \bar{r}_t, \hat{r}^{MC}_{t+1|t} - \bar{r}_t)}{\mathrm{var}_t(\hat{r}^{MC}_{t+1|t} - \bar{r}_t)}. \quad (9)$$

Empirically, δ^* is estimated straightforwardly from (9) by replacing the population ratio with that of the sample covariance to the sample variance. Let $\hat{\delta}^{MC}$ be the estimate. When all data are used, this will yield the in-sample iterated mean combination (IMC) forecast,

$$\hat{r}^{IMC}_{t+1|t} = (1 - \hat{\delta}^{MC})\bar{r}_t + \hat{\delta}^{MC}\hat{r}^{MC}_{t+1|t}. \quad (10)$$

Now replacing $\hat{r}^{MC}_{t+1|t}$ by $\hat{r}^{WC}_{t+1|t}$, the same procedure produces $\hat{\delta}^{WC}$. Then we obtain the in-sample IWC forecast,

$$\hat{r}^{IWC}_{t+1|t} = (1 - \hat{\delta}^{WC})\,\bar{r}_t + \hat{\delta}^{WC}\hat{r}^{WC}_{t+1|t}. \quad (11)$$

Obviously, any other forecasts may also be used in the iteration, but, for simplicity, we consider only $\hat{r}^{IMC}_{t+1|t}$, $\hat{r}^{IWC}_{t+1|t}$, $\hat{r}^{IMD}_{t+1|t}$, and $\hat{r}^{ITC}_{t+1|t}$ in the remainder of this paper.

To generate out-of-sample forecasts, at time t we use the data up to t to estimate the delta, so that only information available at t is used to forecast the return at $t+1$. In this way, the forecast will not contain any future data and thus will be out-of-sample. This is the recursive procedure that we use to obtain out-of-sample forecasts.

To see how the iterated combination works, we examine a simple example. This example mainly serves the purpose of providing an intuition for the working of iterated combination, despite the fact that the optimal solution of the mean squared error problem should yield an improved forecast in general. Suppose that the true return obeys the following process:

$$r_{t+1} = 3\% + 0.02z_{1t} + \cdots + 0.02z_{nt} + \varepsilon_{t+1}, \quad (12)$$

where $z_{1t}, \ldots, z_{nt}$ are known predictors that are independently distributed cross-sectionally with mean 0 and variance 1, and the residual has 0 mean as usual. Since the predictors have equal distributions, the MC and WC will be the same. Ignoring the estimation errors, we have

$$\hat{r}^{WC}_{t+1|t} = 3\% + 0.02 \times \frac{z_{1t} + \cdots + z_{nt}}{n}. \quad (13)$$

Then

$$r_{t+1} - \hat{r}^{WC}_{t+1|t} = \frac{n-1}{n} 0.02(z_{1t} + \cdots + z_{nt}) + \varepsilon_{t+1}, \quad (14)$$

and hence its mean squared error is

$$E_t[r_{t+1} - \hat{r}^{WC}_{t+1|t}]^2 = \frac{(n-1)^2}{n} \times 0.02^2 + \sigma^2_{\varepsilon}, \quad (15)$$

where σ^2_{ε} is the variance of ε_t. As shown, the number of predictors affects the mean squared error almost linearly. The greater the number of predictors, the greater the error. On the other hand, from (12), the optimal forecast is

$$\hat{r}^*_{t+1|t} = 3\% + 0.02 \times (z_{1t} + \cdots + z_{nt}), \quad (16)$$

whose mean squared error is

$$E_t[r_{t+1} - \hat{r}^*_{t+1|t}]^2 = \sigma^2_{\varepsilon},$$

which is smaller and independent of n. Note that for an appropriate large sample size T, the sample mean should be close to 3%. In other words,

$$\begin{aligned} \hat{r}^*_{t+1|t} &\approx \bar{r}_t + 0.02 \times (z_{1t} + \cdots + z_{nt}) \\ &= (1-n)\bar{r}_t + n\hat{r}^{WC}_{t+1|t}. \end{aligned} \quad (17)$$

This shows that in the special case of independently and identically distributed predictors, δ should be large and close to n. In practice, since the predictors have complex joint distributions, δ can, of course, be quite different from n. The main point is that it is generally not between 0 and 1 in contrast to many portfolio problems.

Finally, it should be pointed out that δ is estimated in applications, and so there are estimation errors that are sample dependent. Therefore, there is no guarantee that the iterated combination forecast will always outperform the sample mean or the underlying combination forecast. This is because, when the sample size is small or the system is highly unstable, the errors in estimating δ can be large. Nevertheless, in our applications, this is not a concern. As will be shown, the iterated combination forecasts always outperform the simple combination forecasts substantially.

1.3. Relation to PLS

Our paper differs from others in the literature by considering a large number of predictors that are related to future stock and bond returns. To maximize the benefits from a wealth of data, we employ an efficient method to extract the relevant information from this large set of predictors to obtain better forecasts. In a separate vein, the PLS forecast method, pioneered by Wold (1966) and further developed by Kelly and Pruitt (2013, 2015), provides another powerful procedure for abstracting information from a large set of predictors. It will be useful to compare our method with the PLS.

Interestingly, of our four combinations, IMC reduces to the PLS in the case of linear regression models. To see this, let $PLS_t = \sum_{i=1}^{N} \omega_i z_{it}$ be the PLS combination of the predictors; then

$$\begin{aligned} r_{t+1} &= a + \beta_{PLS} PLS_t + v_t \\ &= a + \omega_1 \beta_{PLS} z_{1t} + \cdots + \omega_N \beta_{PLS} z_{Nt} + v_t. \end{aligned} \tag{18}$$

Comparing this with (2), we have

$$\frac{\beta_i}{\beta_j} = \frac{\omega_i}{\omega_j} = \frac{\operatorname{cov}(r_{t+1}, z_{it})}{\operatorname{cov}(r_{t+1}, z_{jt})}. \tag{19}$$

On the other hand, it is straightforward to verify that the above equality also holds true for the IMC if the individual forecasts are obtained from univariate linear predictive regressions.

Therefore, when we use IMC in linear regression models, we are effectively using the PLS. However, IMC can be applied to both linear and nonlinear models. Moreover, IWC is different from the PLS, and, as will be shown later, it generally performs better than the IMC in our applications.

1.4. Out-of-Sample Performance Measures

We conduct extensive out-of-sample analysis in addition to common in-sample studies (e.g., Greenwood and Hanson 2013) to establish firmly the predictability of corporate bond returns. The out-of-sample forecast is exactly the same as the in-sample forecast, except that it is done recursively. That is, if the out-of-sample forecast evaluation begins from time m, we use all available data or information up to time $t = m - k$ to estimate the parameters of the predictive model to construct the forecast of the excess return k periods ahead, at time $t + k = m$, where k is the forecast horizon. This recursive forecast procedure applies to any future time until $T - k$. Following Campbell and Thompson (2008) and Pettenuzzo et al. (2014), we impose the economic restriction that the risk premium must be positive to be consistent with theory. From an econometric standpoint, the sign restriction can minimize the impact of volatile out-of-sample forecasts when a regression is estimated over a short sample period. However, we note that our results are robust to this restriction.

Following the convention in return forecasting (Fama and French 1989, Campbell and Thompson 2008), we evaluate the out-of-sample performance of the model relative to the updated historical average using the out-of-sample R^2 statistic:[5]

$$R_{OS}^2 = 1 - \frac{\sum_{q=m}^{T-k} (r_{q+k} - \hat{r}_{q+k})^2}{\sum_{q=m}^{T-k} (r_{q+k} - \bar{r}_{q+k})^2}, \tag{20}$$

where r_{q+k} is the realized return at $q+k$, $\hat{r}_{q+k}$ ($\bar{r}_{q+k}$) is the out-of-sample forecast from the predictive regression model (historical average), q is the time that the forecast is made, k denotes the periods ahead in the forecast, and T is the sample size. The out-of-sample R^2 gauges the improvement of the predictive regression model over the historical average forecast in terms of mean squared prediction errors (MSPEs). When $R_{OS}^2 > 0$, the predictive regression forecast performs better than the historical average forecast. We test the statistical significance of R_{OS}^2 by the p-value of the MSPE-adjusted statistic of Clark and West (2007), following the procedure in Rapach et al. (2010).[6] For the forecast horizon longer than a month, we use the Hodrick (1992) method to account for the effect of overlapping residuals on standard errors.[7]

Moreover, to assess whether adding variables significantly improves the predictive power of the model, we employ the test of Harvey, Leybourne, and Newbold (Harvey et al. 1998; hereafter referred to as HLN). The null hypothesis is that the model 1 forecast encompasses the model 2 forecast, against the one-sided alternative that the former does not encompass the latter. Let $e_{t+k} = (\hat{u}_{1,t+k} - \hat{u}_{2,t+k})\hat{u}_{1,t+k}$, where $\hat{u}_{1,t+k} = r_{t+k} - \hat{r}_{t+k}^{M1}$, $\hat{u}_{1,t+k} = r_{t+k} - \hat{r}_{t+k}^{M2}$, and $\hat{r}_{t+k}^{M1}$ and $\hat{r}_{t+k}^{M2}$ are the k-period-ahead return forecasts by models 1 and 2, respectively. The test statistic is

$$HLN = \frac{T - m - k - 1}{T - m - k} [\hat{V}(\bar{e})^{-1/2}] \bar{e},$$

where $\bar{e} = (1/(T-k-m)) \sum_{t=m}^{T-k} e_{t+k}$, $\hat{V}(\bar{e}) = (T-k-m)^{-2} \cdot \sum_{t=m}^{T-k} (e_{t+k} - \bar{e})^2$; HLN has a t distribution with $T - m - k - 1$ degrees of freedom. When the HLN statistic is greater than the critical quantile of the t distribution, the null hypothesis is rejected. This test statistic is used to assess if a set of forecasting variables contains additional information not already in another set of forecasting variables. We use this method to test whether a predictive model encompasses another predictive model. If the forecast of a model is encompassed by another model, we say the latter has more predictive power than the former.

Following Campbell and Thompson (2008), we measure the economic significance of return forecasts. The measure is based on realized utility gains for a mean-variance investor who switches from ignoring predictability to using the predicted return calculated from the out-of-sample forecast. The investor who forecasts the corporate bond return using a model i will allocate a proportion of the portfolio $w_{i,t} = (1/\gamma)(\hat{r}_{t+1|t,i}/\hat{\sigma}^2_{t+1|t})$ to risky bonds at time t, where γ is the risk aversion coefficient, and $\hat{\sigma}^2_{t+1|t}$ is the estimate of the variance of bond excess returns. The realized utility gain of the investor over the out-of-sample period is $\hat{v}_i = \hat{\mu}_i - 0.5\gamma\hat{\sigma}^2_i$, where $\hat{\mu}_i$ and $\hat{\sigma}^2_i$ are the sample mean and variance of returns of the portfolio formed using the bond return forecast of model i. Note that $\hat{v}_i - \hat{v}_0$ gives a direct measure of economic significance between the portfolio choices using forecast model i and benchmark model 0. In our empirical analysis, we use the historical average forecast as the benchmark model. The variance is estimated by the return data in the last five years. We primarily use the variance of a broad or rating portfolio to calculate the utility gain but also examine the robustness of results to the consideration of covariance between bonds of different ratings and maturities. The utility gain measure can also be interpreted as the fee investors would be willing to pay to obtain the forecast instead of using the historical average. A utility gain of 2% or more by the predictive model is usually considered to be economically significant.

2. The Data

Corporate bond data are collected from several sources: the Lehman Brothers Fixed Income (LBFI) database, Datastream, the National Association of Insurance Commissioners (NAIC) database, the Trade Reporting and Compliance Engine (TRACE) database, and Mergent's Fixed Investment Securities Database (FISD). Using individual bond data to form portfolios, we examine return predictability for bonds with different ratings, maturities, and other bond characteristics.

The LBFI database covers monthly data for corporate bond issues from January 1973 to March 1998. The data include month-end prices, accrued interest, rating, issue date, maturity, and other bond characteristics. Datastream reports the daily corporate bond price averaged across all dealers for that bond. We choose U.S. dollar-denominated bonds with regular coupons and obtain the data up to June 2012. The TRACE and NAIC databases contain transaction data for corporate bonds. TRACE coverage begins in July 2002, and NAIC data start from January 1994. TRACE initially covers only a subset of corporate bonds traded in the over-the-counter market, and we supplement it by NAIC, which covers transactions primarily by insurance companies.[8] FISD provides issue- and issuer-specific data such as coupon rate, issue date, maturity date, issue amount, rating, provisions, and other bond characteristics. We merge price data from all sources. Month-end prices are used to calculate monthly returns. The monthly corporate bond log return as of time t is as follows:

$$R_t = \log\frac{(P_t + AI_t) + C_t}{P_{t-1} + AI_{t-1}}, \tag{21}$$

where P_t is the price, AI_t is accrued interest, and C_t is the coupon payment, if any, in month t. We discard the Datastream data if returns are available from other sources, and we choose transaction-based data whenever these data are available. We exclude bonds with maturity less than two years and longer than 30 years and choose only straight bonds to evade confounding effects of embedded options. The sample period runs from January 1973 to June 2012.[9]

We form bond portfolios by rating and maturity. To construct monthly returns of portfolios, we calculate value-weighted mean returns of bonds in each portfolio. In each month, we sort all bonds independently into five rating portfolios and four maturity portfolios using the cutoff points of 5, 7, and 10 years, resulting in 20 portfolios at the intersection of rating and maturity. The short-maturity portfolio is constructed using the bonds with maturity less than 5 years, while the long-maturity portfolio is constructed using the bonds with maturity more than 10 years.

From the literature of equity return forecasts and bond return prediction literature, we consider 27 variables as predictors. We divide predictive variables into three groups: stock market, Treasury market, and corporate bond market variables. The stock market variables include those predictors used in the equity return studies and liquidity indices constructed from stock transaction data. The Treasury bond market variables include those variables which have been shown to have predictive power for Treasury bond returns and the liquidity measures for this market. Finally, the corporate bond market variables include default yield spreads, default return spreads, the issuance quality index, and the debt maturity index. Previous studies have shown that these predictive variables are closely related to credit risk premia. Using different market variables in the regression allows us to see the role of each variable in the predictability of corporate bond returns as a whole and for bonds with different ratings and maturities. In particular, these variables include the following:[10]

1. *Stock market variables*: dividend-price ratio (D/P), dividend yields (D/Y), the earnings-price ratio (E/P),

the dividend-payout ratio (D/E), stock variance (*SVAR*), the book-to-market ratio (B/M), net equity expansion (*NTIS*), S&P 500 index return (*S&P500*), aggregate leverage ratios (*LEV1* and *LEV2*), effective cost (*EC*), Pastor–Stambaugh stock liquidity (*PSS*), and Amihud stock liquidity (*AmS*).

2. Treasury market variables: Treasury bill rate (*TBL*), long-term yield (*LTY*), long-term return (*LTR*), term spread (*TMS*), inflation rate (*INFL*), Cochrane-Piazzesi (2005; hereafter referred to as CP) 5-year factor (*CP5*), Cochrane–Piazzesi (2005) 10-year factor (*CP10*), percentage changes in the money market mutual fund flow (ΔMMMF), and on-/off-the-run spread (*Onoff*). Note that the 5- and 10-year CP forward factors are computed in real time, not based on the full sample. We only use the available data up to the time of forecast to estimate the CP factors for forecasting future returns, and so there is no lookahead bias.

3. *Corporate bond market variables*: default yield spread (*DFY*), default return spread (*DFR*), issuance quality index (*IQ*), debt maturity index (*DM*), and portfolio yield spread (*PYS*).

Using the above-mentioned predictors (27 in total), we consider the following predictive regressions:

a. the predictive regressions using the above individual predictors;

b. the predictive regressions using the combination and iterated combination predictors from the 27 individual predictors;

c. the predictive regression using the principal components of all individual predictors;

d. the multiple predictive regression using term and default spreads as in Fama and French (1989), and then adding Treasury bill rates, lagged high-yield bond returns, and the issuer quality factor as in Greenwood and Hanson (2013). In the extended analysis, we also run multiple regressions with all predictors and subsets of predictors and compare their performance with that of the iterated combination forecast.[11]

3. Empirical Results

3.1. In-Sample Predictability

To understand the role of individual variables in return prediction, we first run regressions of future returns of corporate bonds with different ratings against each predictor. All monthly forecasts are based on monthly nonoverlapping value-weighted bond returns, and quarterly forecasts are based on overlapping bond returns where quarterly returns is the sum of current and past two monthly returns. Returns are all based on log returns. The unreported results (see the Internet appendix) show that individual predictors have varying predictive power and each of these predictors contains information in different dimensions for returns of bonds with distinct quality and premium components (e.g., default and liquidity). This finding suggests that there is considerable room for combining individual forecasts to increase the predictive power of the model. The individual forecasts can be combined using the traditional methods such as mean, median, trimmed mean, and weighted average combination methods. However, as we demonstrated earlier, the iterated combination method can substantially improve the performance of the predictive model. We next investigate this possibility based on the in-sample and out-of-sample results of forecasts for corporate bond returns.

The left panel of Table 1 reports the results of in-sample predictions by using typical combination methods and our new iterated combination methods. Besides the MC and the WC, we consider the median combination (MD) and trimmed mean combination (TC). Consistent with the literature, combination forecasts are valuable in combining the information. Furthermore, the MC and WC appear to perform the best among the four combination methods.

Better than expected, iterated combination methods further improve drastically the already impressive in-sample combination forecasts. As shown in the right panel of the table, each of the four iterated combination forecasts has a substantially higher R^2 than its respective combination counterpart. For example, the in-sample R^2 of IWC for AAA bonds (a value-weighted portfolio of AAA bonds across maturities) is 9.46, which is 4.6 times that of the WC. However, the IMD has much lower R^2 than that of the IWC, indicating that the relative performance of the iterated combination forecasts is linked to the strength of the underlying combination methods. Consistent with Kelly and Pruitt (2013), the IMC, which is equivalent here to the PLS, is a powerful predictor. Nevertheless, the IWC improves even further and provides overall the best forecasts. All of the above results are robust to different ratings and maturities.[12]

We now compare the IWC forecast with three major alternative forecasts in the literature. The first is the PCA forecast that is based on the first principal component of all predictors. The second is the FF model (Fama and French 1989) model that uses default spreads and term spreads as predictors. The third is the Greenwood–Hanson (2013) model (hereafter referred to as GH) that uses the Treasury bill rates, lagged high-yield bond returns, and the issuer quality ratio as additional predictors.

The right panel of Table 1 compares the in-sample R^2 values for different models. The results cover both the rating portfolios as well as the maturity portfolios in each rating category. The FF model performs well with an average in-sample R^2 of 4% for the monthly forecast and 8.52% for the quarterly forecast. Though not reported in the table, the R^2 values are 7.07% and 13.02% over 1973–1987, which covers part of the FF sample period, and 2.01% and 4.97%

Table 1. In-Sample *R*-Squares

	Combination forecast vs. iterated combination forecast								Other predictors and comparison								
	Combination forecast (%)				Iterated combination forecast (%)				Monthly (%)				Quarterly (%)				
	MC	MD	TC	WC	IMC	IMD	ITC	IWC	PCA	FF	GH	Δ	IWC	PCA	FF	GH	Δ
All																	
AAA	2.06	0.99	1.79	2.07	9.28	5.27	8.49	9.46	0.20	2.06	0.73	7.40	11.73	0.41	4.85	3.49	6.88
AA	3.04	1.64	2.77	3.04	11.18	6.57	10.55	11.30	0.44	4.48	2.72	6.82	17.13	0.86	9.40	6.88	7.73
A	2.93	1.27	2.62	2.93	11.33	9.13	10.72	11.46	0.07	5.28	3.59	6.18	16.59	0.11	10.38	8.53	6.21
BBB	3.71	1.73	3.41	3.70	13.40	8.34	12.82	13.52	0.26	7.25	5.04	6.27	19.44	0.56	14.72	11.38	4.72
Junk	3.95	2.12	3.53	3.90	14.03	11.06	13.04	14.11	0.17	6.33	4.73	7.78	21.17	0.36	12.04	9.51	9.13
All	2.76	1.25	2.44	2.79	11.01	5.72	10.12	11.22	0.25	4.00	2.35	7.22	15.64	0.47	8.52	6.34	7.12
Short (2 years < Maturities < 5 years)																	
AAA	2.46	0.96	2.08	2.53	12.38	9.42	11.38	12.72	0.01	2.28	1.11	10.44	13.71	0.01	4.42	4.04	9.29
AA	3.54	1.65	3.17	3.56	14.14	8.81	13.41	14.40	0.25	4.99	3.58	9.41	20.29	0.37	9.76	8.00	10.53
A	3.41	1.30	3.03	3.41	14.22	11.49	13.58	14.41	0.00	5.99	4.78	8.42	19.27	0.01	10.33	9.28	8.94
BBB	3.86	1.94	3.52	3.84	15.00	11.20	14.47	15.20	0.05	7.62	5.88	7.58	20.35	0.05	13.81	10.90	6.54
Junk	3.49	2.03	3.10	3.43	12.96	10.68	12.24	13.02	0.11	5.80	4.32	7.22	18.32	0.17	8.99	6.83	9.33
All	3.12	1.22	2.71	3.18	13.92	8.59	12.77	14.31	0.07	4.27	2.77	10.04	17.75	0.11	7.85	6.18	9.90
5 years < Maturities < 7 years																	
AAA	2.43	0.86	2.02	2.47	11.93	6.44	10.47	12.21	0.02	1.73	0.78	10.48	13.34	0.01	3.78	3.17	9.56
AA	3.03	1.59	2.72	3.03	11.44	6.66	10.61	11.65	0.38	3.89	2.22	7.76	16.04	0.71	7.94	5.59	8.10
A	3.23	1.37	2.86	3.22	12.67	9.78	11.84	12.85	0.06	5.30	3.94	7.55	16.91	0.06	9.36	7.93	7.55
BBB	3.06	1.36	2.78	3.06	11.73	6.47	11.02	11.84	0.17	5.65	3.97	6.19	17.67	0.49	12.78	9.88	4.89
Junk	2.74	1.26	2.46	2.72	10.21	6.64	9.52	10.32	0.06	4.16	2.62	6.16	19.89	0.13	12.76	10.75	7.13
All	2.72	1.17	2.39	2.75	11.30	5.95	10.33	11.61	0.16	3.76	2.37	7.85	15.68	0.39	8.10	6.34	7.58
7 years < Maturities < 10 years																	
AAA	1.75	0.79	1.48	1.76	8.08	3.85	7.25	8.25	0.13	1.37	0.23	6.88	9.71	0.74	4.12	3.20	5.59
AA	2.79	1.42	2.52	2.79	10.43	5.66	9.70	10.57	0.35	3.81	2.18	6.76	15.38	0.71	8.24	6.23	7.14
A	2.73	1.07	2.42	2.73	10.63	7.55	9.89	10.74	0.05	5.05	3.35	5.69	15.80	0.06	10.60	8.92	5.20
BBB	3.23	1.38	2.90	3.24	12.03	9.34	11.39	12.15	0.06	7.24	5.36	4.91	18.39	0.10	14.88	12.45	3.51
Junk	3.35	1.61	3.04	3.34	10.92	6.41	10.00	11.00	0.58	4.38	3.18	6.62	18.37	0.81	10.77	8.90	7.60
All	2.51	1.01	2.21	2.52	9.95	4.69	9.15	10.12	0.15	3.83	2.30	6.29	15.00	0.38	8.66	6.78	6.34
Long (Maturities > 10 years)																	
AAA	1.62	0.65	1.42	1.61	7.27	3.74	6.61	7.33	0.16	1.99	0.63	5.34	10.95	0.81	6.71	4.94	4.24
AA	2.38	1.32	2.19	2.35	7.91	4.47	7.47	7.94	0.64	3.61	1.79	4.33	14.26	1.43	8.47	5.83	5.79
A	2.24	0.92	1.97	2.23	7.83	4.73	7.03	7.86	0.20	4.51	2.68	3.35	14.71	0.60	11.18	8.39	3.53
BBB	2.88	1.42	2.68	2.87	9.81	5.52	9.42	9.90	0.55	5.45	3.75	4.45	16.88	1.63	13.47	10.97	3.41
Junk	2.66	1.02	2.26	2.63	9.83	5.08	8.68	9.77	0.01	4.79	2.88	4.98	19.18	0.00	10.49	8.31	8.69
All	2.42	0.95	2.11	2.43	9.28	3.81	8.35	9.39	0.26	3.60	1.83	5.79	15.25	0.87	10.01	7.42	5.24
Average	2.87	1.31	2.55	2.87	11.20	7.10	10.41	11.35	0.20	4.48	2.92	6.87	16.49	0.45	9.58	7.58	6.91

Notes. The left panel reports the in-sample *R*-squares of combination and iterated combination forecasts, including the MC, the MD, the TC, the WC, and their iterated combination forecast for the portfolio (All) that includes all bonds and portfolios by rating and maturity (Maturity). The right panel reports the in-sample *R*-squares of the PCA, the FF model, and the GH model at the monthly horizon and for the four models including IWC at the quarterly horizon. The difference between in-sample *R*-squares of the IWC and FF models is denoted by Δ. The sample period is from January 1973 to June 2012.

in the post-FF period 1988–2012, for monthly and quarterly forecast horizons, respectively. Although the predictability degenerates somewhat since the publication of their paper, the FF variables do have significant predictive power over time.

Surprisingly, the GH model does not perform better than FF even in the in-sample forecast, even though it has more predictors. This finding echoes previous studies on stock predictability that show adding more variables will not necessarily improve forecasting performance (see Welch and Goyal 2008). The reason is that, econometrically, the predictive multiple regression model tends to perform poorly with highly correlated regressors.

The principal component predictor PCA has poor performance too. All its R^2 values are substantially below 1%. These results are based on the first factor of PCA. For robustness, we also use multiple factors extracted by the principal component method. Unreported results show that in-sample R^2 values increase as the number of factors increases to three (e.g., by about four percentage points with three factors), but the IWC still outperforms substantially.

The last column (Δ) for each forecast horizon in the right panel reports the difference in the R^2 values between the prediction using the IWC predictor and that using the FF model, to further highlight the improvement of the IWC. The differences are all positive, with the maximum value equal to 10.48% for the monthly forecast and 10.53% for the quarterly forecast. The superior performance of the IWC is robust across ratings and maturities. The results suggest that relying on the FF model will substantially underestimate the true predictability and that there is value of using a large set of predictors.

3.2. Out-of-Sample Predictability

Welch and Goyal (2008) argue forcefully that in-sample predictability can be due to overfitting, and out-of-sample forecasting is a more stringent test of return predictability. Henceforth, we shall focus on out-of-sample results in the remaining analysis.

The left panel of Table 2 reports out-of-sample R^2 values of the four forecasting combination methods and their iterated analogues. There are several major findings. First, all of the combination methods deliver positive and statistically significant R^2 values, implying that they are indeed robust forecasting procedures that are able to predict returns both in-sample and out-of-sample. Second, the MD (median combination

Table 2. Out-of-Sample R-Squares

	Combination forecast vs. iterated combination forecast								Other predictors and comparison								
	Combination forecast (%)				Iterated combination forecast (%)				Monthly (%)				Quarterly (%)				
	MC	MD	TC	WC	IMC	IMD	ITC	IWC	PCA	FF	GH	Δ	IWC	PCA	FF	GH	Δ
All																	
AAA	2.18[a]	1.10[a]	1.82[a]	2.21[a]	5.21[a]	5.60[a]	4.52[a]	5.32[a]	−0.47	1.60[a]	0.71[b]	3.72	5.73[b]	−1.67	3.37[a]	1.05	2.36
AA	3.46[a]	1.84[a]	3.02[a]	3.49[a]	9.88[a]	8.58[a]	9.09[a]	9.93[a]	0.15	5.52[a]	2.79[a]	4.41	15.60[a]	−0.58	10.41[a]	5.63[b]	5.19
A	2.63[a]	1.14[a]	2.29[a]	2.67[a]	8.55[a]	7.55[a]	8.12[a]	8.56[a]	−1.38	4.86[a]	1.40[b]	3.70	11.30[a]	−3.08	9.24[a]	4.40[c]	2.06
BBB	3.00[a]	1.24[a]	2.70[a]	3.05[a]	9.58[a]	6.71[a]	9.63[a]	9.68[a]	−0.70	5.99[a]	1.51[b]	3.69	15.27[a]	−2.06	13.42[a]	6.56[c]	1.85
Junk	2.97[a]	1.11[a]	2.51[a]	3.03[a]	11.22[a]	7.72[a]	10.94[a]	11.34[a]	−0.68	5.72[a]	3.03[a]	5.62	16.98[a]	−2.54	10.31[a]	3.91[b]	6.67
All	2.92[a]	1.44[a]	2.55[a]	2.96[a]	7.70[a]	6.78[a]	7.10[a]	7.82[a]	−1.32	3.58[a]	0.72[b]	4.24	12.07[a]	−2.83	7.28[a]	3.11[b]	4.79
Short (2 years < Maturities<5 years)																	
AAA	2.80[a]	1.16[a]	2.23[a]	2.87[a]	4.46[a]	5.01[a]	3.34[a]	4.67[a]	−2.32	1.14[a]	0.41[b]	3.53	5.70[a]	−5.04	2.44[b]	0.93	3.26
AA	4.03[a]	1.92[a]	3.43[a]	4.09[a]	11.88[a]	11.11[a]	10.70[a]	11.98[a]	−1.19	5.94[a]	3.25[a]	6.04	17.68[a]	−3.74	10.46[a]	6.00[b]	7.22
A	3.02[a]	1.28[a]	2.55[a]	3.07[a]	10.81[a]	6.75[a]	10.11[a]	10.83[a]	−2.22	5.47[a]	2.65[b]	5.36	13.08[b]	−4.85	9.16[a]	5.54[c]	3.92
BBB	3.12[a]	1.51[a]	2.78[a]	3.18[a]	11.24[a]	9.95[a]	11.31[a]	11.24[a]	−1.37	6.39[a]	2.14[c]	4.85	14.96[b]	−4.18	12.14[b]	4.34[c]	2.82
Junk	2.32[a]	0.92[a]	1.93[a]	2.37[a]	8.91[a]	5.17[a]	8.87[a]	8.86[a]	−0.55	5.42[a]	3.39[b]	3.44	10.32[b]	−1.64	7.45[a]	2.32[b]	2.87
All	3.15[a]	1.37[a]	2.68[a]	3.23[a]	7.19[a]	5.20[a]	6.41[a]	7.41[a]	−2.78	2.98[a]	0.08[b]	4.43	11.20[a]	−6.68	5.41[a]	2.13[b]	5.79
5 years < Maturities < 7 years																	
AAA	1.86[a]	0.99[a]	1.62[a]	1.89[a]	5.10[a]	5.10[a]	4.69[a]	5.25[a]	−0.69	0.55[c]	−0.54	4.70	5.35[c]	−2.65	0.82[a]	−1.99	4.53
AA	3.08[a]	1.54[a]	2.68[a]	3.12[a]	8.69[a]	7.34[a]	7.96[a]	8.70[a]	0.55	4.91[a]	2.55[b]	3.79	13.77[a]	0.34	9.14[b]	4.76	4.63
A	2.61[a]	1.06[a]	2.20[a]	2.64[a]	8.40[a]	7.19[a]	8.17[a]	8.38[a]	−1.01	5.03[a]	2.32[c]	3.35	11.36[b]	−2.58	8.71[a]	4.81	2.65
BBB	2.34[a]	0.96[a]	2.15[a]	2.37[a]	6.48[a]	3.53[a]	6.62[a]	6.49[a]	0.06	3.80[a]	0.15	2.69	12.94[a]	−0.54	10.22[a]	3.65	2.72
Junk	2.09[a]	0.74[a]	1.81[a]	2.12[a]	6.27[a]	4.38[a]	6.46[a]	6.28[a]	−0.54	3.13[a]	0.63[c]	3.15	12.73[a]	−1.55	6.87[a]	1.81[c]	5.86
All	2.67[a]	1.22[a]	2.3[a]	2.72[a]	6.90[a]	5.94[a]	6.17[a]	7.05[a]	−1.32	3.03[a]	0.52[b]	4.02	12.76[a]	−3.33	7.18[a]	3.36[c]	5.58
7 years < Maturities < 10 years																	
AAA	1.88[a]	0.94[a]	1.55[a]	1.91[a]	3.16[a]	3.82[a]	2.39[a]	3.27[a]	−0.71	0.99[c]	0.16	2.28	3.77[b]	−0.57	3.07[a]	2.18[b]	0.70
AA	3.03[a]	1.64[a]	2.67[a]	3.06[a]	8.22[a]	8.48[a]	7.75[a]	8.26[a]	0.15[c]	4.82[a]	2.06[b]	3.44	12.75[a]	−0.61	9.12[a]	4.32[c]	3.63
A	2.58[a]	1.06[a]	2.28[a]	2.61[a]	7.42[a]	6.39[a]	7.25[a]	7.44[a]	−1.11	5.06[a]	1.25[c]	2.38	10.81[a]	−2.60	9.70[a]	4.59[c]	1.11
BBB	3.09[a]	1.18[a]	2.75[a]	3.15[a]	8.14[a]	6.56[a]	8.22[a]	8.29[a]	−0.85	7.3[a]	2.30[b]	0.99	15.09[a]	−2.12	15.5[a]	8.86[b]	−0.41
Junk	3.08[a]	1.25[a]	2.70[a]	3.13[a]	8.86[a]	5.51[a]	8.78[a]	9.02[a]	0.35	3.81[a]	0.72[b]	5.21	13.10[a]	−1.24	9.93[a]	3.13[a]	3.17
All	2.78[a]	1.16[a]	2.43[a]	2.82[a]	7.09[a]	5.54[a]	6.65[a]	7.17[a]	−1.19	4.17[a]	1.35[b]	3.00	13.00[a]	−2.89	9.38[a]	3.46[a]	3.62
Long (Maturities > 10 years)																	
AAA	1.36[a]	0.63[a]	1.18[a]	1.37[a]	2.11[a]	2.02[a]	2.37[a]	2.10[a]	−0.28	1.49[b]	−0.39	0.61	3.72[b]	0.83	5.93[a]	−0.34	−2.21
AA	2.38[a]	1.21[a]	2.10[a]	2.39[a]	5.66[a]	5.31[a]	5.45[a]	5.72[a]	0.71[b]	4.11[a]	1.71[b]	1.61	11.58[a]	0.85	8.58[a]	2.99[c]	3.00
A	2.03[a]	0.87[a]	1.82[a]	2.04[a]	4.88[a]	5.08[a]	4.74[a]	4.92[a]	−0.27	4.00[a]	0.91[b]	0.92	10.37[a]	−0.70	10.12[a]	3.77[b]	0.25
BBB	2.49[a]	1.08[a]	2.31[a]	2.51[a]	5.77[a]	4.74[a]	5.96[a]	5.77[a]	0.87[b]	4.04[a]	1.76[b]	1.73	12.49[a]	2.76[b]	9.29[a]	5.41[b]	3.20
Junk	1.82[a]	0.59[a]	1.48[a]	1.86[a]	6.59[a]	3.23[a]	6.24[a]	6.78[a]	−0.97	3.7[a]	0.35[b]	3.08	12.30[a]	−2.72	7.11[a]	0.24[b]	5.19
All	2.29[a]	1.00[a]	2.01[a]	2.33[a]	5.41[a]	4.04[a]	4.99[a]	5.49[a]	−1.12	2.99[a]	−0.02	2.50	10.83[a]	−2.22	8.80[a]	2.35[c]	2.03
Average	2.63	1.17	2.28	2.67	7.39	6.01	7.03	7.47	−0.74	4.05	1.33	3.42	11.62	−2.01	8.35	3.44	3.27

Notes. The left panel reports the monthly out-of-sample R-squares of the four combination forecasts and their iterated combination forecasts for the portfolio (All) that includes all bonds and portfolios by rating and maturity (Maturity). The right panel reports the out-of-sample R-squares of the PCA, the FF model, and the GH model at the monthly horizon, and the four models including the IWC at the quarterly horizon. The p-value is based on the MSPE-adjusted statistic of Clark and West (2007). The difference between out-of-sample R-squares of the IWC and FF forecasts is denoted by Δ. The sample period is from January 1973 to June 2012, while the out-of-sample forecast starts from January 1983.

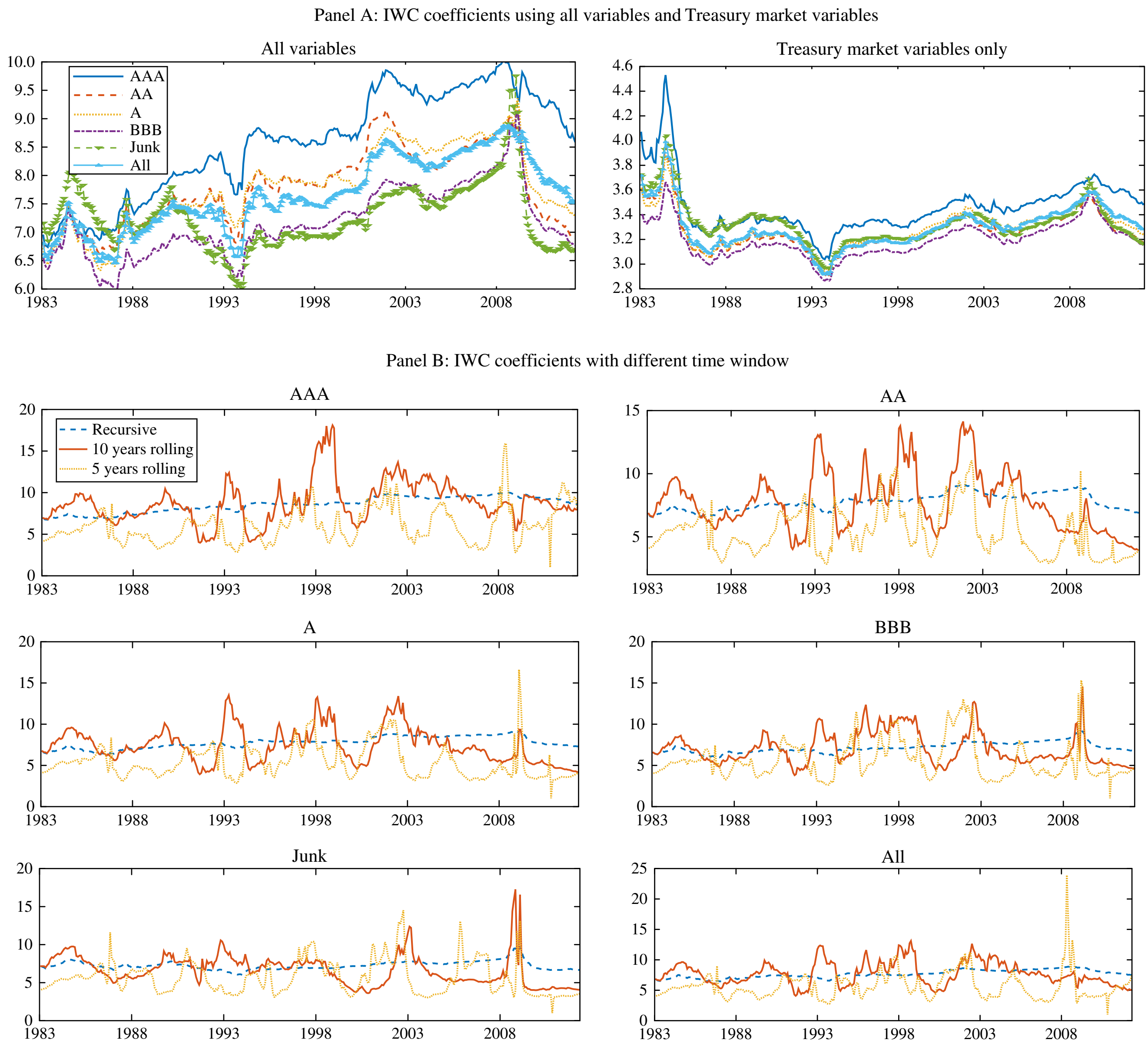
[a], [b], and [c] denote the significance levels of 1%, 5%, and 10%, respectively.

forecast) seems to have the worst performance among the four combinations, suggesting that the forecasts across individual predictors are asymmetric. Third, the iterated combinations improve their original combinations substantially. Overall, results strongly suggest that iterated combination is a superior method to generate more efficient forecasts.

Figure 1 plots the time series of δ estimates from the combination forecast regression. The left graph of panel A plots δ estimates using all predictors and the right graph of panel A plots estimates using only the Treasury market predictors for comparison. Results show that the value of δ is much larger than 1. Also, when more predictors are used (see the left panel), the δ value is higher, which is consistent with our analysis in Section 1.2.

In addition, to examine the stability of δ estimates, we estimate this parameter using different sliding windows. Panel B of Figure 1 plots δ's using the sliding windows of 5 and 10 years of past monthly observations as well as the recursive rolling window that uses all information up to forecasting time t. Results

Figure 1. (Color online) IWC Coefficients



Notes. Panel A plots the IWC coefficients using all variables and Treasury market variables. The left panel plots the coefficients if all variables are used. The right panel plots the coefficients when only Treasury market variables are used. Panel B plots the time series of the coefficient of iterated weighted combination forecast ($\hat{\delta}$) for bond portfolios using different sliding windows of historical data. The curve of "recursive" uses all historical data available at time t, "10-year rolling" uses the data of the last 10 years at time t, and "5-year rolling" uses the last five years of data.

show that the current recursive rolling method produces much smoother estimates for δ's than the 5-year and 10-year rolling windows.

We next compare the out-of-sample performance of the IWC with the other three predictive models: PCA, FF, and GH. Note that when performing the out-of-sample forecast at time t, we only use the available information up to time t to perform forecasts. Hence, the PCA method uses available information from all predictors only up to t, and the FF and GH are based on recursive regressions.

The right panel of Table 2 compares the out-of-sample R^2 of the four predictive regression models. Similar to the in-sample results in Table 1, FF and GH have sizable out-of-sample predictive ability. The out-of-sample R^2 for all bonds using the FF model is 3.58% for the monthly forecast and 7.28% for the quarterly forecast. The results for the GH model are weaker but still significant. Most of the out-of-sample R^2 values of PCA are negative.

The IWC has the best out-of-sample predictive performance among the four models (see the last column of the left panel and the first column of quarterly results in the right panel). All out-of-sample R^2 values are significantly positive. For the monthly forecast, it can be as high as 11.98% (AA short-maturity portfolio). The average out-of-sample R^2 of the IWC is 7.82% for all bonds at the monthly horizon. For the quarterly forecast, the highest R^2 is 17.68% (AA short-maturity portfolio), and the R^2 is 12.07% for all bonds. Both are much higher than the out-of-sample R^2 values of FF and GH. Interestingly, these out-of-sample R^2 values are substantially higher than those for forecasting the stock risk premium. For example, Rapach et al. (2010) report an out-of-sample R^2 of only about 1% for the quarterly forecast during 1975–2005. Thus, the results suggest that the corporate bond market is much more predictable than the stock market.

The last column (Δ) for each forecast horizon in the right panel reports the differences in out-of-sample R^2 values between the IWC and FF. All the differences are overwhelmingly positive, indicating that the IWC model has much higher predictive power than the FF. The improvement of monthly forecasts by the IWC is greater than that of quarterly forecasts. Similar to in-sample results, the improvement is quite robust across ratings and maturities, and it is attributable to a better use of the information in a large set of predictors by the iterated combination method. We also conduct forecast encompassing tests using the HLN statistics of Harvey et al. (1998) to formally evaluate which model is better. The empirical results (see the Internet appendix) show that the IWC model encompasses other models strongly.

3.3. Economic Significance

Table 3 reports results of economic significance measured by utility gains or certainty equivalent returns (CER). Certainty equivalent gains are all annualized values based on monthly or quarterly forecasts. The risk aversion coefficient is set equal to 5, and the optimal weight is between 0 (short-sales constraint) and 5, similar to other studies such as Thornton and Valente (2012) and Goh et al. (2013). The left panel reports the results of monthly forecasts, while the right panel reports quarterly forecasts. As in the case for the out-of-sample R^2 values, we compare the four models: IWC, PCA, FF, and GH.

Consistent with the out-of-sample R^2 values, the utility gains of the IWC are much larger than those of the FF, which in turn are often much larger than those of the GH and PCA. The utility gains of the IWC are all positive except for only one case for junk bonds with long maturity. Even in this particular case, the IWC still performs the best among the four models at the monthly forecast horizon. Across rating and maturity portfolios, the utility gains of the FF model are mostly economically insignificant. Similar to the results based on R^2 values, the GH model performs worse than the FF model, while the PCA is the worst performer whose gains are all negative except for BBB bonds with long maturity.

The last column in both panels of Table 3 reports the differences in utility gains between the IWC and FF models. These differences are overwhelmingly positive for both monthly and quarterly forecast horizons. The improvement in economic value by the IWC is greater for the monthly forecast, and results are again robust across ratings and maturities.

Overall, results show that the gains of the out-of-sample forecasts by the IWC are not only statistically significant as shown earlier, but also economically meaningful. For the monthly forecast, the utility gain is 5.74% for the sample that includes all bonds. For the quarterly forecast, the gain is 3.77% for all bonds. The utility gains of the IWC are much larger than other models and also considerably higher than those for the stock market reported by Rapach et al. (2010), suggesting there is substantial economic value of using a large set of predictors and the proposed methodology in bond return forecasts.

3.4. Multiple PCA Predictors

Another issue is that up until now, we have used only the first PCA factor as a predictor in out-of-sample forecasts. This raises a potential concern that we may have underestimated the predictive power of the PCA model, as additional factors may contain useful information. To address this concern, we rerun the out-of-sample regressions using two to five factors extracted from PCA. Table 4 reports the out-of-sample forecast results using more PCA factors. For brevity, we only report the results associated with three and five factors. The predictive power increases somewhat

Table 3. Utility Gains

Maturity	Rating	Monthly (%) IWC	PCA	FF	GH	ΔU	Quarterly (%) IWC	PCA	FF	GH	ΔU
All	AAA	6.10	−0.53	1.37	0.16	4.73	2.46	−0.74	1.75	0.01	0.72
	AA	6.04	−0.28	2.33	−0.10	3.71	4.25	−0.53	2.89	1.65	1.35
	A	5.94	−1.86	1.74	−1.11	4.20	3.04	−1.79	1.47	1.11	1.57
	BBB	6.15	−1.73	1.31	−2.56	4.84	1.84	−1.73	0.82	−2.21	1.03
	Junk	2.35	−1.73	−1.77	−3.69	4.12	1.13	−1.95	0.55	0.50	0.58
	All	5.74	−1.41	1.58	−0.38	4.16	3.77	−1.46	1.86	0.37	1.91
Short (2 years < Maturity < 5 years)	AAA	2.28	−2.90	−0.77	−0.38	3.05	0.69	−2.24	0.17	−0.67	0.51
	AA	4.32	−1.67	0.51	−0.35	3.81	3.46	−1.46	1.81	0.79	1.66
	A	3.54	−3.06	0.62	−1.02	2.92	2.27	−2.87	0.36	0.30	1.91
	BBB	6.90	−2.36	1.14	−1.70	5.76	1.12	−3.12	−0.61	−3.35	1.73
	Junk	4.28	−0.85	2.94	1.23	1.34	3.73	−0.76	3.10	2.28	0.62
	All	2.86	−3.16	−0.85	−1.48	3.71	1.42	−2.81	−0.41	−1.06	1.83
5 years < Maturity < 7 years	AAA	5.15	−1.43	−1.58	−1.63	6.73	0.48	−1.41	−1.60	−0.80	2.08
	AA	5.78	−0.10	1.75	−0.98	4.03	3.54	−0.31	2.41	1.41	1.12
	A	3.28	−1.79	−0.53	−2.80	3.81	0.96	−1.67	0.43	1.40	0.53
	BBB	4.81	−0.95	−1.56	−4.75	6.37	0.24	−1.14	−0.97	−4.01	1.20
	Junk	2.23	−0.74	−1.11	−4.70	3.34	2.78	−0.98	1.08	−1.34	1.70
	All	4.97	−1.73	1.05	−0.69	3.92	3.23	−1.51	1.30	−0.08	1.94
7 years < Maturity < 10 years	AAA	5.11	−0.94	−0.36	−0.56	5.47	2.15	−0.29	1.09	0.87	1.06
	AA	7.66	−0.32	2.11	−0.98	5.55	3.84	−0.58	2.42	1.23	1.42
	A	5.74	−1.56	−0.29	−3.47	6.03	2.54	−1.47	0.53	1.06	2.01
	BBB	6.49	−1.44	2.73	−0.83	3.76	2.53	−1.46	1.52	−1.02	1.00
	Junk	5.78	−0.22	−1.72	−7.12	7.50	1.22	−0.94	−0.60	−5.19	1.82
	All	6.93	−1.37	2.06	−0.43	4.87	4.31	−1.16	2.56	1.04	1.75
Long (Maturity > 10 years)	AAA	2.82	−0.67	−0.51	−2.30	3.33	0.91	−0.23	1.55	−0.36	−0.64
	AA	4.00	−0.06	1.46	−1.75	2.54	2.25	−0.28	2.39	1.26	−0.15
	A	2.80	−0.92	−0.82	−3.00	3.62	0.55	−0.95	1.18	0.34	−0.63
	BBB	3.60	0.15	−0.27	−4.11	3.87	−0.07	0.43	1.57	−2.00	−1.64
	Junk	−0.60	−0.78	−4.35	−0.70	3.75	1.85	−1.00	−2.70	2.22	4.55
	All	6.42	−1.43	1.33	−2.54	5.09	3.87	−1.08	2.24	0.14	1.63
Average		4.65	−1.26	0.32	−1.82	4.33	2.21	−1.25	1.01	−0.14	1.21

Notes. This table reports the annualized utility gains of the IWC, the PCA, the FF model, and the GH model for the portfolio (All) that includes all bonds and portfolios by rating and maturity (Maturity). The difference between the utility gains of the IWC and FF models is denoted by ΔU. The sample period is from January 1973 to June 2012, while the out-of-sample forecast starts from January 1983.

when the number of factors increases from 1 to 3. For example, for monthly (quarterly) out-of-sample forecasts, the average R^2 value increases to 1.75% (3.21%) when the number of factors increases from 1 to 3. However, the R^2 value declines after the number of factors exceeds 3 and becomes negative with five factors. For example, the out-of-sample R^2 for the monthly (quarterly) forecast is 1.17% (−1.42%) when the number of factors used in forecasts is equal to 5. A similar pattern is found for the utility gain, but results are worse for quarterly forecasts. Results show that an increase in the number of factors improves the forecast power of PCA only modestly and up to a certain limit. Despite the increase in the forecast power of PCA when including more factors, the IWC continues to outperform PCA by a substantial margin of about 6% and 9% in out-of-sample R^2 values for monthly and quarterly horizons, respectively (see Table 2). These findings again suggest that the iterated combination produces the best out-of-sample forecasts.

3.5. Longer-Horizon Forecasts

We have shown thus far that corporate bond returns are highly predictable at monthly and quarterly horizons. We next examine the forecasts for longer horizons. For brevity, we provide the results only for junk bonds, as the premium of these bonds is particularly interesting. Panel A of Table 5 reports return forecasts at longer horizons ranging from two quarters to one year for speculative-grade bonds. Results show that junk bond returns are predictable over longer horizons. For the in-sample forecasts, the IWC model continues to perform much better than the Fama–French model. The improvement in in-sample R^2 by the IWC over the FF model is quite substantial and increases with the forecast horizon. The increases in R^2 range from 15.3% to 20.5% for the whole sample that includes all bonds. Similarly, the out-of-sample forecasts show predictability at longer horizons. The IWC model consistently outperforms the FF model across all horizons. The out-of-sample R^2 values of the IWC are quite high,

Table 4. Forecasts with Multifactor PCA Models

		Monthly				Quarterly			
		R^2_{OS} (%)		Utility gains (%)		R^2_{OS} (%)		Utility gains (%)	
	Maturity	PCA(3)	PCA(5)	PCA(3)	PCA(5)	PCA(3)	PCA(5)	PCA(3)	PCA(5)
All	AAA	0.87[b]	0.21[b]	1.03	5.11	−2.34	−6.22	−1.69	−0.83
	AA	6.04[a]	2.62[a]	3.54	5.02	10.38[a]	5.41[b]	0.07	0.52
	A	8.60[a]	0.81[a]	4.01	4.01	13.11[a]	3.53[b]	0.14	0.16
	BBB	12.36[a]	1.94[a]	3.32	5.27	17.23[a]	9.57[a]	0.02	0.30
	Junk	10.59[a]	4.76[a]	3.12	−0.46	17.21[a]	9.28[a]	−2.31	−2.73
	All	1.75[a]	1.17[a]	2.31	4.86	3.21[a]	−1.42	−2.10	−1.55
Short	AAA	0.71[a]	−2.16	−1.59	2.39	−4.26	−8.27	−3.57	−2.70
(2 years < Maturity < 5 years)	AA	7.54[a]	6.07[b]	0.51	2.17	11.87[a]	12.00[b]	−0.97	−0.57
	A	7.12[a]	0.62[a]	0.98	1.70	9.98[a]	4.76[b]	−0.65	−0.56
	BBB	12.19[a]	1.95[a]	3.93	5.41	18.86[a]	5.94[a]	−1.84	−2.15
	Junk	10.55[a]	2.65[a]	4.87	0.61	14.07[b]	4.61[b]	−0.62	−1.08
	All	3.54[a]	0.65[b]	−0.78	1.74	2.13[a]	−1.20	−4.02	−3.25
5 years < Maturity < 7 years	AAA	0.60[b]	0.43[b]	0.39	6.46	−6.75	−13.13	−0.98	−0.08
	AA	4.77[a]	0.57[b]	3.25	4.00	8.96[a]	2.04[c]	0.35	0.76
	A	8.73[a]	5.22[a]	2.85	2.86	9.08[b]	4.05[c]	0.12	0.23
	BBB	10.66[a]	1.97[a]	3.91	3.87	14.18[a]	6.18[b]	0.51	−0.17
	Junk	6.44[a]	1.87[a]	2.43	1.08	12.52[a]	5.13[b]	−0.29	−0.46
	All	1.70[a]	−0.63	2.18	4.18	4.49[a]	−1.13	−1.31	−0.73
7 years < Maturity < 10 years	AAA	−0.58	−0.97	0.40	4.06	1.93[a]	−1.88	−0.74	0.00
	AA	4.18[a]	1.06[b]	3.48	4.29	8.10[a]	3.91[c]	−0.03	0.48
	A	5.87[a]	−2.81[b]	3.92	3.39	11.38[a]	1.29[c]	0.27	0.07
	BBB	9.65[a]	2.98[a]	5.45	5.79	19.89[a]	13.40[a]	1.09	0.88
	Junk	7.50[a]	4.66[a]	4.21	3.02	14.68[a]	9.28[a]	−0.60	−1.20
	All	1.57[a]	−0.44	2.12	4.56	5.10[a]	0.24[c]	−1.38	−0.71
Long	AAA	−2.79	−3.03	0.08	2.43	−2.36	−8.05	−1.38	−0.46
(Maturity > 10 years)	AA	5.96[a]	−0.95	1.37	0.95	7.56[b]	2.39[b]	−0.24	0.20
	A	3.18[b]	−0.67	0.70	0.62	11.15[a]	3.36[b]	−0.12	−0.28
	BBB	8.07[a]	3.22[a]	4.80	3.02	16.14[a]	8.60[a]	2.24	2.15
	Junk	8.00[a]	0.76[a]	−0.25	−1.65	10.80[a]	3.75[a]	−2.38	−3.44
	All	0.16[b]	−1.29	2.24	4.82	2.04[a]	−3.11	−1.82	−1.08
Average		5.52	1.11	2.29	3.19	8.68	2.48	−0.81	−0.61

Notes. This table reports the out-of-sample R-squares (R^2_{OS}) and utility gains of multifactor PCA models for the portfolio that includes all bonds (All) and portfolios by rating and maturity (Maturity). PCA(3) and PCA(5) are the principal component models with three and five factors, respectively. The p-value of R^2_{OS} is based on the MSPE-adjusted statistic of Clark and West (2007).

[a], [b], and [c] denote the significance levels of 1%, 5%, and 10%, respectively.

ranging from 23.5% to 28.1% from two quarters to one-year horizon.

The predictability of junk bond returns in longer horizons is of economic significance. As shown in panel A of Table 5, the utility gains from using the IWC model are overwhelmingly positive. For the whole sample including all bonds, the IWC model delivers higher economic value than the FF model by a margin of 2.87%–3.20% in terms of CER. Results show that the economic value of using the IWC predictor is significant.

3.6. Joint Asset Allocations Across Ratings and Maturities

The analysis of economic gains in the preceding sections is carried out separately for each bond portfolio. This approach has been used in equity studies (see, for example, Campbell and Thompson 2008) and can be viewed as a way to improve an asset allocation. Once the allocation is given to an asset, this analysis shows the gains of an allocation based on predictability versus the one based on historical estimates. More recently, in studying the portfolio allocation for Treasury bonds, Thornton and Valente (2012) and Sarno et al. (2016) carry out the asset allocation jointly for Treasury bonds of all maturities. These studies deal with an asset allocation problem consisting of bonds with different maturities. In this section, we extend our analysis of economic gains by using a similar approach to implement asset allocation, which considers multiple risk bonds jointly across all maturities. In our context of corporate bonds with various ratings, we carry out the asset allocation jointly for bonds with different ratings and maturities, respectively. Value-weighted bond portfolios in the rating/maturity buckets are used to calculate the joint asset allocation.

Table 5. Extended Tests

Panel A: Longer-horizon forecasts of junk bonds

	Two quarters (%)			Three quarters (%)			One year (%)		
Maturity	FF	IWC	Δ	FF	IWC	Δ	FF	IWC	Δ
				In-sample R-squares					
All	19.65	34.96	15.31	26.63	44.93	18.29	30.78	51.27	20.49
Short	17.20	33.81	16.61	22.69	41.97	19.28	26.35	48.50	22.15
5 years < Maturity < 7 years	20.50	33.95	13.45	27.04	43.37	16.33	29.73	49.28	19.55
7 years < Maturity < 10 years	15.55	28.64	13.09	21.09	37.32	16.23	24.91	43.91	19.00
Long	17.34	31.27	13.93	24.49	41.42	16.94	29.68	48.74	19.05
				Out-of-sample R-squares					
All	15.88[a]	23.51[a]	7.63	21.73[a]	24.78[b]	3.05	26.46[a]	28.05[b]	1.59
Short	12.83[a]	18.42[b]	5.58	17.25[a]	19.96[b]	2.71	20.76[a]	23.13[b]	2.37
5 years < Maturity < 7 years	12.24[a]	19.23[a]	6.99	20.47[a]	23.72[b]	3.25	27.73[a]	29.18[b]	1.45
7 years < Maturity < 10 years	13.76[a]	20.31[a]	6.55	20.32[a]	24.53[b]	4.21	25.07[a]	28.17[b]	3.10
Long	11.75[a]	17.08[a]	5.32	17.67[a]	20.54[a]	2.87	21.46[a]	24.89[b]	3.43
				Utility gains					
All	−2.84	0.03	2.87	−2.48	0.72	3.20	−2.06	0.81	2.87
Short	0.52	3.55	3.03	0.45	2.81	2.36	0.63	2.51	1.87
5 years < Maturity < 7 years	−0.51	2.62	3.13	1.27	2.90	1.63	2.67	2.61	−0.06
7 years < Maturity < 10 years	−1.26	−0.38	0.88	−1.69	−0.44	1.25	−0.19	0.27	0.46
Long	−2.52	2.08	4.60	−1.00	2.69	3.69	−0.16	2.74	2.89

Panel B: Utility gains of joint asset allocation

	Monthly (%)			Quarterly (%)		
	FF	IWC	Δ	FF	IWC	Δ
Rating portfolio						
Short (2 years < Maturity < 5 years)	1.00	4.57	3.56	0.88	2.98	2.10
5 years < Maturity < 7 years	−0.09	6.53	6.62	0.45	3.95	3.49
7 years < Maturity < 10 years	2.21	9.59	7.38	2.99	4.92	1.93
Long (Maturity > 10 years)	0.92	3.20	2.28	0.00	3.19	3.18
All	1.16	7.52	6.37	1.73	4.14	2.40
Average	1.04	6.28	5.24	1.21	3.84	2.62
Maturity portfolio						
AAA	−2.15	2.05	4.20	0.34	1.46	1.11
AA	0.22	4.52	4.31	1.88	3.51	1.64
A	0.39	2.56	2.17	1.31	2.88	1.57
BBB	0.39	3.05	2.66	1.63	1.68	0.04
Junk	2.25	2.59	0.34	2.78	3.69	0.91
All	−0.14	3.70	3.84	0.63	2.64	2.01
Average	0.16	3.08	2.92	1.43	2.64	1.21

Notes. This table reports the results of extended tests including longer-horizon forecasts and utility gains of joint asset allocation. Panel A reports the results of longer-horizon forecasts over two quarters, three quarters, and one year for junk bonds. The difference between the results of the IWC and FF models is denoted by Δ. Panel B reports the annualized utility gains by allocating different rating and maturity (Maturity) portfolios jointly, where Δ reports the difference between the utility gains of the IWC and FF results.

Panel B of Table 5 reports the results based on joint asset allocations of portfolios across either ratings or maturities. For brevity, we compare only the results for the FF and IWC models, as these two models are our focus. The upper panel reports the results of joint asset allocations across ratings for each maturity group and for all maturities (All), while the lower panel reports results by considering correlations of risk premia across maturities for bonds in each rating category and for all rated bonds (All).

The results show that the utility gains remain significant with the joint allocation approach. Across ratings, the lowest utility gain is 3.2% for the long-maturity group at the monthly horizon, which is economically significant. The differences in the utility gains based on the forecasts of the IWC and FF models are sizable. Across maturities, utility gains are smaller but still of significant economic importance. The lowest utility gain is 2.05% for AAA bonds at monthly horizon. Interesting, even in this case, the gain is still above 2%. Overall, the results show that our economic gain analysis is robust to the consideration of joint asset allocations of multiple bonds across all maturities or ratings, suggesting again that the IWC delivers superior forecasting performance.[13]

3.7. Transaction Costs

The preceding analysis shows that the gains from using the predictive model are economically significant. This implies that investors will be better off by taking an investment strategy based on model forecasts. However, the gains of using return forecasts of the predictive models are overstated by ignoring transaction costs. In this subsection, we explore the impacts of transaction costs and examine whether returns and utility gains based on the model forecasts survive transaction costs. We first compute the turnover rates of portfolios each month. We then estimate transaction costs by accounting for the turnover rates of the portfolios formed by the forecasts of different models, and we report returns and utility gains net of transaction costs.

The left side of panel A in Table 6 reports mean returns of investment portfolios based on the forecasts of random walk (RW) or historical average, FF and the IWC models. Results show that the investment strategy based on the IWC model on average produces the highest returns for the portfolio that considers all bonds (All) and across portfolios of different ratings and maturities. For example, average monthly returns for the portfolio that includes all bonds (All) are 1.84% when using the IWC model as opposed to 1.49% for the FF model and 1.11% for the random walk model. Results show that it is more profitable for bond investors to use the IWC model than the FF or RW model to forecast returns.

The turnover rates are reported in the middle of panel A for the portfolios formed by the three forecasts. Results show that the turnover rate is the highest for the portfolios using the iterated combination forecasts.

Table 6. Transaction Costs and Economic Significance

Panel A: Turnover ratios and breakeven transaction costs

	Return (%)			Turnover ratio			Breakeven cost (%)		
	RW	FF	IWC	RW	FF	IWC	1	2	3
All									
AAA	0.77	1.10	1.43	0.04	0.22	1.06	1.87	0.65	0.40
AA	0.87	1.39	1.72	0.05	0.25	0.99	2.55	0.90	0.45
A	0.91	1.48	1.75	0.06	0.27	0.93	2.68	0.97	0.42
BBB	1.07	1.88	2.22	0.06	0.31	0.95	3.18	1.29	0.53
Junk	1.60	2.33	3.04	0.09	0.33	1.03	2.95	1.53	1.02
All	1.11	1.49	1.84	0.07	0.27	0.99	1.88	0.80	0.49
Short (2 years < Maturity < 5 years)									
AAA	1.18	1.08	1.33	0.10	0.28	1.28	−0.55	0.13	0.25
AA	1.17	1.24	1.62	0.09	0.30	1.14	0.34	0.43	0.46
A	1.23	1.33	1.53	0.12	0.32	1.03	0.48	0.33	0.28
BBB	1.30	1.74	2.09	0.11	0.34	1.07	1.95	0.82	0.47
Junk	1.18	1.81	2.29	0.08	0.32	1.06	2.66	1.14	0.65
All	1.38	1.30	1.60	0.11	0.32	1.14	−0.38	0.22	0.37
5 years < Maturity < 7 years									
AAA	0.70	0.96	1.48	0.03	0.20	0.97	1.61	0.83	0.67
AA	0.71	1.24	1.60	0.03	0.23	1.06	2.64	0.88	0.44
A	0.87	1.43	1.67	0.06	0.25	0.89	2.88	0.96	0.37
BBB	0.90	1.49	1.83	0.06	0.26	0.88	2.85	1.13	0.55
Junk	0.95	1.49	2.16	0.04	0.23	0.85	2.72	1.48	1.09
All	1.02	1.33	1.66	0.06	0.25	1.00	1.62	0.68	0.44
7 years < Maturity < 10 years									
AAA	0.77	1.05	1.45	0.04	0.18	1.05	1.93	0.67	0.46
AA	0.76	1.42	1.88	0.03	0.24	0.96	3.21	1.22	0.65
A	0.82	1.40	1.77	0.04	0.24	0.83	2.94	1.20	0.62
BBB	0.97	1.88	2.07	0.05	0.27	0.74	4.07	1.58	0.39
Junk	1.00	1.68	2.63	0.03	0.23	0.82	3.51	2.07	1.60
All	1.01	1.54	1.93	0.05	0.23	0.87	2.85	1.11	0.60
Long (Maturity > 10 years)									
AAA	0.68	1.03	1.38	0.02	0.14	0.70	2.91	1.03	0.62
AA	0.74	1.38	1.84	0.02	0.20	0.77	3.69	1.47	0.79
A	0.73	1.45	1.70	0.03	0.23	0.67	3.60	1.50	0.56
BBB	0.76	1.70	2.01	0.02	0.23	0.67	4.63	1.92	0.70
Junk	2.00	2.64	3.32	0.10	0.27	0.66	3.58	2.33	1.75
All	0.99	1.64	2.09	0.03	0.24	0.88	3.20	1.31	0.71
Average	1.00	1.50	1.90	0.06	0.26	0.93	2.47	1.09	0.63

Table 6. (Continued)

Panel B: Utility gains net of transaction costs						
	Monthly (%)			Quarterly (%)		
	FF	IWC	Δ	FF	IWC	Δ
All						
AAA	−1.05	3.84	4.88	0.86	2.06	1.20
AA	0.28	2.56	2.28	2.00	3.40	1.41
A	1.52	4.33	2.81	1.01	2.73	1.72
BBB	1.69	2.87	1.18	0.14	1.00	0.85
Junk	−0.39	−1.14	−0.75	0.28	0.28	0.00
All	0.54	3.54	3.00	1.20	3.29	2.09
Short (2 years < Maturity < 5 years)						
AAA	1.77	7.29	5.52	0.11	1.76	1.65
AA	2.12	5.39	3.26	1.54	3.58	2.04
A	3.03	7.51	4.48	0.74	3.29	2.55
BBB	5.17	8.19	3.01	−0.25	1.68	1.92
Junk	3.47	2.31	−1.16	2.96	3.20	0.24
All	2.03	6.31	4.28	0.18	2.44	2.26
5 years < Maturity < 7 years						
AAA	−2.29	3.50	5.79	−2.10	0.50	2.60
AA	−0.99	1.64	2.64	1.65	2.70	1.05
A	0.54	3.08	2.54	0.09	0.87	0.78
BBB	0.78	2.11	1.34	−1.65	−0.57	1.07
Junk	−1.38	−1.49	−0.10	0.44	1.66	1.22
All	0.09	3.51	3.41	0.66	2.80	2.14
7 years < Maturity < 10 years						
AAA	−1.61	3.16	4.77	0.33	1.31	0.98
AA	−0.31	3.14	3.45	1.64	3.00	1.36
A	−0.27	3.60	3.86	0.03	2.23	2.20
BBB	2.54	3.61	1.08	0.73	1.96	1.23
Junk	−0.95	−0.09	0.87	−0.86	0.32	1.18
All	0.85	4.42	3.57	1.71	3.73	2.02
Long (Maturity > 10 years)						
AAA	−0.76	0.48	1.24	0.84	0.13	−0.70
AA	−0.32	−0.71	−0.39	1.79	1.28	−0.51
A	−2.49	−0.97	1.52	0.53	−0.31	−0.84
BBB	−0.94	−1.31	−0.38	0.48	−1.55	−2.03
Junk	−1.14	−2.42	−1.28	−2.85	1.49	4.34
All	0.50	2.93	2.42	1.41	2.97	1.55
Average	0.40	2.71	2.30	0.52	1.77	1.25

Notes. The left part of panel A reports mean returns, and the middle part reports mean turnover ratios of the portfolios using forecasts of the RW, FF, and IWC models, respectively. The right part reports the breakeven cost that makes the return performance of two portfolios indifferent. Breakeven cost 1 is the cost that makes the return of the portfolio using the FF model indifferent from that using the RW model. Breakeven cost 2 (respectively, 3) is the cost that makes the portfolio return using the IWC model indifferent from that using the RW (respectively, FF) model. Panel B reports the utility gains of the IWC and FF models net of transaction costs. Transaction costs per dollar of trading are 0.25% before July 2007 and 0.60% afterward for investment-grade bonds, and 0.35% before July 2007 and 0.50% afterward for junk bonds. The difference of utility gains between the FF and IWC models is Δ. Results are reported for all bonds (All) and portfolios by rating and maturity (Maturity).

This pattern holds for the portfolio that includes all bonds and for the portfolios formed by rating and maturity. As the IWC model incorporates more variables and uses the information more efficiently in return forecasts, it is not surprising to see the portfolios based on this forecast have higher turnover rates than those based on other forecasts.

We next calculate the breakeven transaction cost that will render investors indifferent between the two competing strategies based on the models (see Grundy and Martin 2001, Thornton and Valente 2012). The transaction cost is set equal to a fixed proportion of the value traded in the different bond portfolios. Following Thornton and Valente (2012), we calculate the breakeven cost of two strategies between strategy 1 and strategy 2 using the following formula:

$$\frac{E(r_{p,1}) - E(r_{p,2})}{TO_1 - TO_2}, \tag{22}$$

where $E(r_{p,1})$ and $E(r_{p,2})$ are the portfolios' mean returns, and TO_1 and TO_2 are the average turnover ratios of strategies 1 and 2, respectively. In comparing a predictive model with the historical average (random walk) or another model, an investor with a transaction cost lower than the breakeven cost will prefer the former model to the latter.

The right side of panel A in Table 6 reports the transaction cost per dollar of trading that will make the return of the portfolio using a forecast model indifferent from that of the portfolio using another forecast model. Column 1 under the breakeven cost section compares portfolios formed by the FF model with those by the RW model, column 2 compares the IWC model with the RW model, and column 3 compares the IWC model with the FF model. For example, for the portfolio that is constructed from all bonds, it will take additional cost of 0.8% for the investment strategy using the IWC model to be indifferent from the strategy using the RW model (see the sixth row in column 2 under breakeven cost). Edwards et al. (2007) report an average transaction cost of about 24 basis points per dollar trading for a median size of corporate bond trade. On the basis of this transaction cost estimate, the profits from the investment strategy using the IWC forecasts survive transaction costs across the board. Transaction cost tends to be higher for low-grade bonds for a fixed trade size. A question is whether investment strategy will be profitable for speculative-grade bonds. Edwards et al. (2007) estimate a transaction cost of about 35 basis points for junk bonds. On the basis of this transaction cost estimate, predictability profits of junk bonds also survive transaction costs by a wide margin.

Our finding implies investors will be better off by investing in portfolios formed by the forecasts of the predictive models. Corporate bond prices and transaction data are disseminated by TRACE, which covers trading of all publicly traded bonds, and basic bond characteristics such as ratings and maturities are readily available from brokers. Investors can form their rating and maturity portfolios using publicly available data to implement the investment strategy based on the model forecasts. This is especially the case for the corporate bond market, which is dominated by institutional investors who are much more sophisticated and have more resources to conduct the analysis and implement the investment strategy. For retail investors, exchange traded funds (ETFs) tracking the bond indexes at different maturities and ratings are widely available with low transaction costs. These investors can apply our method to forecast returns and invest in ETFs with low trading costs without having to form rating and maturity portfolios by themselves. Thus, our analysis has practical implications for investors to improve their investment performance.

As for stocks, the bond returns tend to be more predictable in bad states of the economy (see the Internet appendix). A related issue is whether returns of low-grade bonds are more predictable because the transaction cost for these bonds is high especially in the bad economy. Jostova et al. (2013) estimate transaction costs of corporate bonds for the period of 2009–2010, which overlaps with the subprime crisis. Their estimates of transaction costs are higher than those reported by Edwards et al. (2007). Jostova et al. (2013) report average transaction cost of 54 basis points for noninvestment grade (NIG) bonds of private firms and 51 basis points for NIG of public firms. Using these high estimates of transaction costs, returns of junk bonds still survive the transaction cost. Furthermore, unreported results (omitted for brevity) show that returns for junk bonds have become more predictable in the post-TRACE period during which the transaction cost is lower than in the pre-TRACE period. Thus, there is little evidence that returns for junk bonds are more predictable because transaction costs for these bonds are higher than other bonds, or bond returns are more predictable in bad states of the economy because of higher transaction costs. However, this argument is subject to a caveat. In the financial crisis, markets for high-risk assets can virtually freeze. Besides trading costs, there is an illiquidity issue. Thus, the profits may not be realizable for high-risk bonds in times of stress when liquidity dries up. This issue requires a further investigation that goes beyond the scope of the current paper.

We next calculate the utility gains net of transaction costs. We first subtract the transaction cost from the portfolio returns and then recalculate the utility gains or certainty equivalent returns. Because our sample period covers the normal and crisis periods, we use the transaction cost estimates of Edwards et al. (2007) for the period before July 2007 and the cost estimates by Jostova et al. (2013) afterward. The use of higher transaction in the post-crisis period is rather conservative. Panel B of Table 6 reports the utility gains net of transaction costs. Results show that utility gains remain positive even after accounting for the transaction costs. The net gains delivered by the IWC are, on average, 3.54% and 3.29% for monthly and quarterly horizons, compared with 0.54% and 1.20% for the FF model. Thus, there is evidence of significant net economic gains from using the predictive models, and the IWC model outperforms the FF model substantially. However, while overall there are net economic gains, results show that transaction cost reduces the gains for junk bonds. For instance, net gains for junk bonds for the monthly horizon become smaller or turn negative after adjusting for transaction costs, although they remain positive for quarterly horizons. Results suggest that transaction costs play a more important role in high-risk bonds.

4. What Drives the Predictive Power?

The analysis above shows that bond and stock market variables contain important information for expected corporate bond returns and the IWC is an effective method for extracting such useful information from these variables. In this section, we examine how the IWC predictor links to economic fundamentals to understand more about the source of its predictive power.

Cochrane (2007) suggests that return forecasts are more plausibly related to macroeconomic risk if the return predictors also demonstrate an ability to forecast business cycle. The predictability can then be more credibly attributed to time-varying risk premia due to changing risks or risk aversion. In what follows, we examine whether the IWC predictor can forecast real economic activity.

Consider the following predictive regression:

$$\Delta Y_{t+1} = \alpha + \beta X_t + \varepsilon_{t+1}, \tag{23}$$

where ΔY_{t+1} is the change in macroeconomic conditions in the next period, and X_t is the IWC predictor for a given bond portfolio in the current period. In this regression, we examine how the IWC predictor is related to the future state of the economy. Since the PCA is widely used for predicting returns, it is of interest to compare the IWC with the PCA in this context. To do so, we simply run similar regressions with X_t replaced by the PCA predictor.

We use eight measures of Y_{t+1} in the predictive regression, including smooth recession probability (*SRP*), industrial production growth (*IPG*), Treasury bill rates (*TBL*), default yield spreads (*DFY*), implied volatility index (*VIX*), expected default frequency (*EDF*), Chicago Fed National Activity Index (*CFNAI*), and Aruoba et al. (2009) business conditions index (*ADSI*).

Table 7 reports results of the predictive regression in (23) for quarterly horizons.[14] The t-values are calculated using the Newey and West (1987) adjusted standard errors. The results strongly indicate that the IWC predictor has high predictive power for the future change in economic conditions. Among all macroeconomic measures, only the results for Treasury bill rates (*TBL*) are not significant. The predictive power of the IWC varies across bond ratings. An important finding is that the IWC predictor associated with lower-grade bonds has much higher predictive power than that associated with higher-grade bonds. For example, when forecasting the *SRP* (recession probability), the adjusted R^2 of the IWC predictor of the BBB bond portfolio is 9.52%, while it is only 3.62% for the IWC predictor of the AAA bond portfolio. The results for other macroeconomic variables show a similar pattern. Results show that the predicted premia by the IWC for BBB and junk bonds have consistently higher predictive power for future economic activity than those of higher-grade bonds. These findings suggest that the expected excess return or risk premium of lower-grade bonds contains substantially more information for future economic activity than that of higher-grade bonds. This evidence supports the prediction of Greenwood and Hanson (2013) that bonds issued by low-quality firms provide more reliable signals for the future economic and financial conditions. Our results are also consistent with the finding of Gilchrist and Zakrajšek (2012) that the excess bond premium has strong predictive power for future economic activity. However, when further examining the predictive power of different rated bond premia, we find that it is the risk premium or credit spread of lower-grade bonds that contains the most predictive information content for future economic activity.

Table 7. Future Macroeconomic Conditions and the Forecasts

X	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)
		$Y = SRP$			$Y = IPG$	
IWC						
AAA	−0.81	−2.28	3.62	1.18	1.37	0.74
AA	−0.75	−3.16	7.13	1.25	2.23	2.15
A	−0.80	−3.69	9.21	1.33	2.63	2.82
BBB	−0.64	−3.72	9.52	1.10	2.75	3.20
Junk	−0.52	−3.44	8.63	0.88	2.52	2.78
PCA	−0.64	−0.38	−0.11	−0.48	−0.12	−0.21
		$Y = TBL$			$Y = DFY$	
IWC						
AAA	−2.12	−1.08	0.75	−0.56	−1.12	0.79
AA	−0.97	−0.83	0.24	−0.67	−1.71	2.93
A	−1.00	−1.02	0.33	−0.70	−1.93	3.78
BBB	−0.52	−0.64	0.02	−0.62	−2.22	4.78
Junk	−0.68	−1.07	0.35	−0.50	−1.95	4.34
PCA	−4.77	−0.50	0.01	0.22	0.11	−0.21
		$Y = VIX$			$Y = EDF$	
IWC						
AAA	−0.29	−1.63	1.97	−0.26	−2.20	3.06
AA	−0.23	−2.38	4.23	−0.20	−2.61	3.92
A	−0.20	−2.43	4.41	−0.21	−2.99	4.87
BBB	−0.16	−2.49	4.31	−0.16	−2.89	4.43
Junk	−0.14	−2.66	5.15	−0.14	−3.07	5.01
PCA	−0.43	−0.52	−0.28	−0.21	−0.48	−0.12
		$Y = CFNAI$			$Y = ADSI$	
IWC						
AAA	2.08	1.93	1.85	2.01	1.53	1.20
AA	2.07	2.88	4.35	2.02	2.48	2.98
A	2.17	3.31	5.51	2.16	2.98	3.94
BBB	1.84	3.62	6.44	1.79	3.20	4.44
Junk	1.42	3.15	5.30	1.46	3.04	4.09
PCA	−0.97	−0.19	−0.19	1.06	0.18	−0.20

Notes. This table reports results of the predictive regression

$$\Delta Y_{t+1} = \alpha + \beta X_t + \varepsilon_{t+1},$$

where ΔY_{t+1} is the change in the recession probability (*SRP*), industrial production growth (*IPG*), Treasury bill rate (*TBL*), default yield spread (*DFY*), implied volatility index (*VIX*), expected default frequency (*EDF*), Chicago Fed National Activity Index (*CFNAI*), or the Aruoba-Diebold-Scotti (2009) business conditions index (*ADSI*); X_t is the IWC predictor for a given rating portfolio or the PCA predictor. The forecast horizon is one quarter. The t-statistics are calculated using the Newey and West (1987) adjusted standard errors. The sample period is from January 1973 to June 2012.

By contrast, none of the results using the PCA to predict future economic conditions is significant and the adjusted R-squares are extremely low, suggesting that the PCA is not a good predictor for the future economic condition. This finding sheds light on the reason why the PCA is a poor predictor for bond risk premia, as shown earlier.

In summary, our empirical results strongly suggest that the predictive power of the IWC forecast is derived from its ability to forecast future macroeconomic conditions. Economic fundamentals are the forces driving time variations in expected corporate bond returns, and the IWC does a good job in tracking the temporal movement of these forces. This explains why the IWC predictor performs much better than the popular PCA factor in predicting bond returns. More importantly, we find that the predominant source of the predictive power of bond premia for business cycle is from low-grade bonds. Empirical evidence shows that low-grade bond premium has the highest forecast power for a wide range of economic indicators.

5. Conclusions

In this paper, we conduct a comprehensive study on the predictability of returns for bonds with different ratings and maturities. We consider a large number of individual predictors, including stock, Treasury, and corporate bond market variables, and we propose a new iterated combination method that combines the information from a large set of predictors. We find that the iterated weighted-average combination forecast performs substantially better than the Fama–French (1989) model, the Greenwood–Hanson (2013) model, the traditional combination forecasts, multiple regression models, and a predictor based on the principal component analysis in in-sample and out-of-sample forecasts in terms of both statistical and economic significance. This finding is robust to bonds with different ratings and maturities, as well as to data based on individual bond or index returns with and without controlling for the return on Treasuries.

Stock and bond market variables contain useful information for predicting corporate bond returns. However, these variables must be carefully combined to preserve the valuable information in them for return forecasts. Improper use of these variables by a naive multiple regression or the principal component analysis destroys the value of these predictors. We show that the proposed iterated combination method is capable of retaining the useful information and reducing noise in individual predictors to obtain much better forecasts. Using this method, we find that the true predictability of corporate bond returns is considerably understated if the predictors are restricted to only a few conventional variables. On source of predictability, we find that risk premia of bonds with different ratings contain important information about future macroeconomic activities at short and long horizons. In particular, the risk premium of high-rated short-maturity bonds provides the information for the short-term prospects, and that of low-rated long-maturity bonds provides the information for the long-term prospects of business conditions.

Acknowledgments

The authors are especially grateful for the insightful comments of two referees, an associate editor, and the department editor (Lauren Cohen). They thank the following for very helpful comments: Michael Brennan, Ivan Brick, Serdar Dinc, Douglas J. Fairhurst, Kenneth French, Yongmiao Hong, George Jiang, Jerry Kallberg, Glenda Kao, Jin-Mo Kim, Cheng F. Lee, Sheen Liu, Frank McIntyre, Mike McNamara, Brian McTier, Daniela Osterrieder, Darius Palia, Ľuboš Pástor, Seth Pruitt, Oded Palmon, Bharat Sarath, Oleg Sokolinskiy, Jack Strauss, Allan Timmermann, Jason Turkiela, Yangru Wu, Fan Yang, Hong Zhang, and Ken Zhong; seminar participants at Monash University, National Taiwan University, Renmin University of China, Rutgers University, Tsinghua University, Washington State University, and Washington University in St. Louis, and Xiamen University; and conference participants at the Q-Group Spring 2014 seminar, the 2014 Annual Missouri Economics Conference, the 2014 New Zealand Finance Colloquium, the 2015 Asian Finance Association Annual Meeting, and the 2015 China International Conference in Finance.

Endnotes

[1] See, for example, Fama and Schwert (1977), Fama and French (1988), Campbell and Shiller (1988), Kothari and Shanken (1997), Pontiff and Schall (1998), Campbell and Vuolteenaho (2004), Ang and Bekaert (2007), Rapach et al. (2010), Henkel et al. (2011), Dangl and Halling (2012), Pettenuzzo et al. (2014), and Rapach et al. (2016) for predicting stocks; and Fama and Bliss (1987), Campbell and Shiller (1991), Cochrane and Piazzesi (2005), Ludvigson and Ng (2009), Almeida et al. (2011), Goh et al. (2013), and Gargano et al. (2016) for predicting government bonds.

[2] In the stock market, Rapach et al. (2010) and Kelly and Pruitt (2013) are examples of using large sets of predictors to obtain significant predictability on the equity risk premium.

[3] Kelly and Pruitt (2013) show that the PLS generates a powerful book-to-market ratio predictor of the stock market, and Huang et al. (2015) find that the PLS provides a strong investor sentiment index for forecasting the equity risk premium.

[4] The unconstrained bivariate regression yields qualitative similar results in our applications below.

[5] To our knowledge, Fama and French (1989) are the first to propose such a statistic, which is used by Welch and Goyal (2008) and Campbell and Thompson (2008) and known subsequently in many predictability studies as Campbell and Thompson out-of-sample R^2.

[6] To perform the test, we first compute the following square error difference: $v_{q+k} = (r_{q+k} - \bar{r}_{q+k})^2 - [(r_{q+k} - \hat{r}_{q+k})^2 - (\bar{r}_{q+k} - \hat{r}_{q+k})^2]$ and then regress v_{q+k} on a constant. The t-statistic of the constant term gives the p-value for the one-sided (upper tail) test.

[7] Correction by the Newey and West (1987) method gives similar results.

[8] The procedure of Bessembinder et al. (2009) is used to filter out canceled, corrected, and commission trades, and daily prices are trade size-weighted average of intraday prices over the day.

[9] In screening the data, we delete the observation with price more than 150 or less than 50 and the data if the time of last available bond price information is more than six months ago. We take the last available price information if there is no price information on the last day of each month and use interpolation to calculate the return.

[10] Information on the 27 predictors is reported in detail in the Internet appendix.

[11] We have also examined other models used by Fama and French (1989) and found similar results, available upon request.

[12] We also compare our two-stage iterative model combination approach with a simpler one-step combination approach where the sample mean $\bar{r}_t$ is added as one of the models in the combination—that is, on the right-hand side of Equations (4) and (5). The results continue to show that our iterative approach significantly outperforms this one-step combination. We thank an anonymous referee for making this suggestion.

[13] Although unreported here, the IWC also improves substantially the existing forecasting procedures, such as the linear regression, PCA, and MC, in forecasting returns on Treasuries. We find that predictability for AAA bonds is a bit higher than that for Treasuries reported in the literature largely because the methodology proposed in this paper produces better forecasts. We also calculate the economic gain for Treasuries using the IWC and find that it is comparable to the result reported by Gargano et al. (2016), who document significant economic value for Treasury return forecasts.

[14] We obtain similar results at monthly and yearly horizons.

References

Almeida C, Graveline JJ, Joslin S (2011) Do interest rate options contain information about excess returns? *J. Econometrics* 164(1):35–44.

Ang A, Bekaert G (2007) Stock return predictability: Is it there? *Rev. Financial Stud.* 20(3):651–707.

Aruoba SB, Diebold FX, Scotti C (2009) Real-time measurement of business conditions. *J. Bus. Econom. Statist.* 27(4):417–427.

Bates JM, Granger CMJ (1969) The combination of forecasts. *Oper. Res. Quart.* 20(4):451–468.

Bessembinder H, Kahle KM, Maxwell WF, Xu D (2009) Measuring abnormal bond performance. *Rev. Financial Stud.* 22(10):4219–4258.

Bhojraj S, Sengupta P (2003) Effect of corporate governance on bond ratings and yields: The role of institutional investors and outside directors. *J. Bus.* 76(3):455–475.

Campbell JY, Shiller RJ (1988) Stock returns, earnings, and expected dividends. *J. Finance* 43(3):661–676.

Campbell JY, Shiller RJ (1991) Yield spreads and interest rates: A bird's eye view. *Rev. Econom. Stud.* 58(3):495–514.

Campbell JY, Thompson SB (2008) Predicting excess stock returns out of sample: Can anything beat the historical average? *Rev. Financial Stud.* 21(4):1510–1531.

Campbell JY, Vuolteenaho T (2004) Inflation illusion and stock prices. *Amer. Econom. Rev.* 94(2):19–23.

Capistrán C, Timmermann A (2009) Forecast combination with entry and exit of experts. *J. Bus. Econom. Statist.* 27(4):428–440.

Choi J, Kim Y (2016) Anomalies and market (dis)integration. Working paper, University of Illinois at Urbana–Champaign, Urbana.

Chordia T, Goyal A, Nozawa Y, Subrahmanyam A, Tong Q (2016) Is the cross-section of expected bond returns influenced by equity return predictors? Working paper, University of California, Los Angeles, Los Angeles.

Clark TE, West KD (2007) Approximately normal tests for equal predictive accuracy in nested models. *J. Econometrics* 138(1):291–311.

Cochrane JH (2007) Financial markets and the real economy. Mehra R, ed. *Handbook of the Equity Premium* (Elsevier, Amsterdam), 237–330.

Cochrane JH, Piazzesi M (2005) Bond risk premia. *Amer. Econom. Rev.* 95(1):138–160.

Dangl T, Halling M (2012) Predictive regressions with time-varying coefficients. *J. Financial Econom.* 106(1):157–181.

Edwards A, Harris LE, Piwowar MS (2007) Corporate bond market transaction costs and transparency. *J. Finance* 62(3):1421–1451.

Fama EF, Bliss RR (1987) The information in long-maturity forward rates. *Amer. Econom. Rev.* 77(4):680–692.

Fama EF, French KR (1988) Dividend yields and expected stock returns. *J. Financial Econom.* 22(1):3–25.

Fama EF, French KR (1989) Business conditions and expected returns on stocks and bonds. *J. Financial Econom.* 25(1):23–49.

Fama EF, Schwert GW (1977) Asset returns and inflation. *J. Financial Econom.* 5(2):115–146.

Gargano A, Pettenuzzo D, Timmermann A (2016) Bond return predictability: Economic value and links to the macroeconomy. Working paper, University of California, San Diego, San Diego.

Gilchrist S, Zakrajšek E (2012) Credit spreads and business cycle fluctuations. *Amer. Econom. Rev.* 102(4):1692–1720.

Goh J, Jiang F, Tu J, Zhou G (2013) Predict bond risk premia using technical indicators. Working paper, Washington University in St. Louis, St. Louis.

Granger CWJ, Ramanathan R (1984) Improved methods of combining forecasts. *J. Forecasting* 3(2):197–204.

Greenwood R, Hanson SG (2013) Issuer quality and corporate bond returns. *Rev. Financial Stud.* 68(6):1483–1525.

Grundy BD, Martin JS (2001) Understanding the nature of the risks and the source of the rewards to momentum investing. *Rev. Financial Stud.* 14(1):29–78.

Harvey DI, Leybourne SJ, Newbold P (1998) Tests for forecast encompassing. *J. Bus. Econom. Statist.* 16(2):254–259.

Henkel SJ, Martin JS, Nardari F (2011) Time-varying short-horizon predictability. *J. Financial Econom.* 99(3):560–580.

Hodrick RJ (1992) Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Rev. Financial Stud.* 5(3):357–386.

Huang D, Jiang F, Tu J, Zhou G (2015) Investor sentiment aligned: A powerful predictor of stock returns. *Rev. Financial Stud.* 28(3):791–837.

Jostova G, Nikolova S, Philipov A, Stahel C (2013) Momentum in corporate bond returns. *Rev. Financial Stud.* 26(7):1649–1693.

Keim DB, Stambaugh RF (1986) Predicting returns in the stock and bond markets. *J. Financial Econom.* 17(2):357–390.

Kelly B, Pruitt S (2013) Market expectations in the cross-section of present values. *J. Finance* 68(5):1721–1757.

Kelly B, Pruitt S (2015) The three-pass regression filter: A new approach to forecasting using many predictors. *J. Econometrics* 186(2):294–316.

Kothari SP, Shanken J (1997) Book-to-market dividend yield, and expected market returns: A time-series analysis. *J. Financial Econom.* 44(2):169–203.

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

Lin H, Wu C, Zhou G (2016) Momentum in corporate bonds. Working paper, Washington University in St. Louis, St. Louis.

Ludvigson SC, Ng S (2009) Macro factors in bond risk premia. *Rev. Financial Stud.* 22(12):5027–5067.

Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3):703–708.

Pettenuzzo D, Timmermann A, Valkanov R (2014) Forecasting stock returns under economic constraints. *J. Financial Econom.* 114(3):517–553.

Pontiff J, Schall LD (1998) Book-to-market ratios as predictors of market returns. *J. Financial Econom.* 49(2):141–160.

Rapach DE, Ringgenberg MC, Zhou G (2016) Short interest and aggregate stock returns. *J. Financial Econom.* 121(1):46–65.

Rapach DE, Strauss JK, Zhou G (2010) Out-of-sample equity premium prediction: Combination forecast and links to the real economy. *Rev. Financial Stud.* 23(2):821–862.

Sarno L, Schneider P, Wagner C (2016) The economic value of predicting bond risk premia. Working paper, City University London, London.

Stock JH, Watson MW (2001) A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Engle RF, White H, eds. *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive Granger* (Oxford University Press, Oxford, UK), 1–44.

Thornton DL, Valente G (2012) Out-of-sample predictions of bond excess returns and forward rates: An asset-allocation perspective. *Rev. Financial Stud.* 25(10):3141–3168.

Timmermann A (2006) Forecast combinations. Elliott G, Granger CWJ, Timmermann A, eds. *Handbook of Economic Forecasting*, Vol. 1 (Elsevier, Amsterdam), 135–196.

Welch I, Goyal A (2008) A comprehensive look at the empirical performance of equity premium prediction. *Rev. Financial Stud.* 21(4):1455–1508.

Wold H (1966) Estimation of principal components and related models by iterative least squares. Krishnaiaah PR, ed. *Multivariate Analysis* (Academic Press, New York), 391–420.