



Forecasting Robust Bond Risk Premia using Technical Indicators

M. Noteboom

414137

Bachelor Thesis Quantitative Finance

Econometrics & Operations Research

Erasmus School of Economics

Supervisor: Xiao Xiao

Second assessor: Xun Gong

Abstract

Extensive research has been done on the relevance of economic variables in predicting bond returns. However, only few studies have considered the importance of technical indicators which are commonly used by investors and traders. This paper studies the predictive power of technical indicators in forecasting bond risk premia. Particularly, this paper assesses the robustness of the indicators by considering several small-sample size problems such as standard error bias and overlapping return bias which frequently appear in predictive regressions for bond returns. For this purpose, I use a bootstrap procedure proposed by Bauer and Hamilton (2018) and a formal test developed by Ibragimov and Müller (2010). I find that technical indicators are not always significant and robust predictors. Only the level and slope of the yield curve consistently show to be significant in forecasting robust bond risk premia. From an asset allocation perspective, a portfolio constructed with the first five principal components of yields and the technical indicators generates the most utility gains.

Keywords: yield curve, bond risk premium predictability, spanning hypothesis, robust methods, technical analysis, principal components

1 Introduction

The predictability of interest rates movements is of great importance for various market participants such as bond investors, policy makers, and financial economists. For bond investors, forecasting interest rates may result in higher bond returns. However, for policy makers understanding the change in future interest rates may help their decision making concerning macroeconomic monetary policy.

Recently, several studies have shown that a variety of macro-economic factors help in predicting excess bond returns (e.g. Cochrane and Piazzesi (2005); Ludvigson and Ng (2009)). However not much extensive research has been done on the relevance of technical indicators in the bond market. Neely et al. (2012) found that technical indicators are significant in predicting the equity risk premium. As the stock and bond markets behave similarly (e.g. Fama and French (1989)), this paper investigates whether technical indicators are also of importance in forecasting bond risk premia.

All the above-mentioned studies make use of predictive regressions for bond returns. However, the particular set-up results in two notable characteristics. The true regressors reflect the information in the yield curve such that they are correlated with the lagged forecast errors. Furthermore, the regressors are highly persistent. Bauer and Hamilton (2018) show that these features result in significant standard error bias which causes the estimated standard errors to be too small in small samples. Moreover, most predictive regressions use monthly data to forecast an annual excess return. The use of overlapping returns leads to correlation in the lagged forecast errors. Hence, conventional standard errors are estimated to be downwards biased. Bauer and Hamilton (2018) show that past methods that have been used to deal with these two econometric issues are not robust as they are subject to significant small-sample and size distortions. Therefore, this paper seeks to answer two questions: First, are technical indicators robust predictors of future bond risk premia? Second, do combinations of technical and macro-economic indicators further improve this predictive power? In addition, I evaluate the economic value of this predictive power for investors by examining the utility gains from an asset allocation perspective.

In this paper I use three types of technical indicators, where the set-up for the first two types of indicators is similar to Goh et al. (2013). The first 48 technical indicators are constructed based on moving averages of lagged forward spreads. The following 15 indicators are formulated based on trading volume data. This is because practitioners regularly use data for the amount of occurred transactions in combination with historical prices to analyze the market behavior. The last 16 indicators are constructed using a momentum strategy. Cochrane and Piazzesi (2005) found that lagged forward rates have significant additional predictive power in forecasting

excess bond returns, which is especially relevant for the forward spread moving average technical indicators used in this paper. However Bauer and Hamilton (2018) show that this result is much weaker in small samples which is one of the main characteristics of macro-economic data. Therefore, I first reproduce the results of Cochrane and Piazzesi (2005) in Bauer and Hamilton (2018) using predictive regressions for excess returns, and show that this evidence is indeed weaker for forecasting robust bond risk premia.

To address the first question, I investigate the so-called spanning hypothesis. According to the spanning hypothesis, the first three principal components (PCs) of the yield curve are adequate to forecast bond returns. These components are commonly indicated as the level, slope and curvature, respectively. Under the null of the spanning hypothesis, technical indicators should add no incremental significant information in forecasting bond returns. Then I use two robust procedures to test this hypothesis. The first is a bootstrap procedure, designed by Bauer and Hamilton (2018), which is constructed specifically to test the null of the spanning hypothesis. In this procedure, the level, slope, and curvature are fit with a VAR model. Various studies, such as Litterman and Scheinkman (1991), have acknowledged that these three factors explain most of the variation in the entire yield curve. Under the bootstrap, it is possible to evaluate the robustness of standard tests by calculating their small-sample size. The second procedure is based on a formal test suggested by Ibragimov and Müller (2010). In this test, the dataset is divided into subsamples such that the coefficients can be estimated separately. Afterwards, a conventional t -test can be applied on the coefficients across the subsamples. Bauer and Hamilton (2018) show that the IM-test successfully deals with the standard error bias which conventional t -tests suffer from.

In order to answer the second question, I use a predictive regression which combines the information in the current yield curve with the constructed technical indicators. In this regression, I use the first five PCs of yields together with the technical indicators as predictors. Afterwards, I investigate the additional predictive power of the indicators using the same procedures as I did to answer the first research question.

Three main conclusions can be drawn from this paper. First, technical indicators are not always significant and robust predictors of future bond risk premia. Second, the addition of macro-economic factors in the form of higher-order PCs of yields does not improve this predictive power. In all findings, the current yield curve reflects all the relevant information for future bond returns in a robust way. Third, the economic value of technical indicators is of considerable size but less than the economic value of the level, slope and curvature. I find that the utility gains for an investor is the highest when he/she uses forecasts based on the information contained in

the first five PCs of yields as well as the technical indicators.

The paper contributes to the literature by evaluating the importance of robust technical indicators in forecasting bond risk premia. Particularly, the paper considers several econometric issues such as standard error bias and overlapping return bias. Both of these biases are very relevant in the context of time series analysis and may lead to different small-sample results for the regressors. Moreover, the paper adds to the literature by evaluating the economic value of bond risk premia forecasts from an asset allocation perspective.

The rest of the paper is structured in the following way. In section 2, I describe the construction of the technical indicators and the methods used to evaluate the predictive power and robustness of these predictors. Section 3 presents the data and shows summary statistics. Robust results of Cochrane and Piazzesi (2005) are shown in section 4. Section 5 and 6 display empirical results of the predictive power of technical indicators and macro-economic factors. In section 7, I assess the economic value of this predictive power. Concluding remarks are provided in section 8.

2 Methodology

2.1 Construction of the technical indicators

I use a similar notation for the excess bond returns and yields as in Goh et al. (2013). Let $p_t^{(n)}$ be the log price of a n -year discount bond at the end of period t . Hence, the log yield of a n -year discount bond at the end of period t is $y_t^{(n)} \equiv -\frac{1}{n}p_t^{(n)}$. The n -year bond price at the end of period t is $f_s^{(n)} \equiv f_t^{(n)} - y_t^{(1)}$, where $f_t^{(n)} \equiv p_t^{(n-1)} - p_t^{(n)}$ is the forward rate at the end of period t for future yields between period $t + n - 1$ and $t + n$.

For the construction of the technical indicators, I consider three types of indicators based on popular technical strategies. The first is a forward spread moving average rule MA^{fs} which states whether one should buy ($S_t = 1$) or sell ($S_t = 0$) at the end of period t by comparing the n -year forward spreads of two moving averages

$$S_t = \begin{cases} 1 & \text{if } MA_{s,t}^{fs,(n)} > MA_{l,t}^{fs,(n)} \\ 0 & \text{if } MA_{s,t}^{fs,(n)} \leq MA_{l,t}^{fs,(n)} \end{cases}, \quad (1)$$

with

$$MA_{j,t}^{fs,(n)} = (1/j) \sum_{k=0}^{j-1} f_{s_{t-k/12}}^{(n)}, \quad \text{for } j = s, l.$$

Here, $f_{s_{t-k/12}}^{(n)}$ is the forward spread for the n -year discount bond at the end of period $t - k/12$, and s (l) is the length of the short (long) forward spread moving average. Then the forward

spread moving average rule at the end of period t with maturity n and length s (l) can be denoted as $MA_{s(l),t}^{fs,(n)}$. The trading signal S_t can be interpreted as follows: if the forward rates for the n -year discount bonds are decreasing relative to the one-year bond yields, then the short forward spread moving average will most likely be lower than the long forward spread moving average. Therefore, an investor would want to sell the bond. The opposite reasoning applies when an investor wants to take a long position in the bond. I analyze the monthly forward spread moving average rule for $n = 2,3,4,5$, $s = 3,6,9$ and $l = 18,24,30,36$.

The second type of technical indicators is constructed based on volume data at the end of period t , OBV_t . However, since volume data for bond trading is not available for the whole sample period, I instead use data for stock market trading volume. Various studies have shown that stock market volume indicators can be a valid proxy for bond market volume indicators (e.g. Campbell and Vuoltenaho (2004)). I define the trading volume-based indicator as

$$OBV_t = \sum_{k=0}^{12t-1} VOL_{t-k/12} D_{t-k/12},$$

where $VOL_{t-k/12}$ measures the trading volume between period $t - (k + 1)/12$ and $t - k/12$, and $D_{t-k/12}$ is a binary variable that is equal to 1 if $P_{t-k/12} - P_{t-(k+1)/12} \geq 0$ and -1 otherwise, with P_t the closing price of the stock index at the end of period t . Afterwards, I define S_t similar to the first type of technical indicators

$$S_t = \begin{cases} 1 & \text{if } MA_{s,t}^{OBV} \leq MA_{l,t}^{OBV} \\ 0 & \text{if } MA_{s,t}^{OBV} > MA_{l,t}^{OBV} \end{cases}, \quad (2)$$

with

$$MA_{j,t}^{OBV} = (1/j) \sum_{k=0}^{j-1} OBV_{t-k/12}, \quad \text{for } j = s, l.$$

Here, $MA_{s(l),t}^{OBV}$ is the volume-based trading rule at the end of period t , where s (l) is the length of the short (long) moving average of the volume-based indicator. This trading signal S_t can be interpreted as follows: a relatively large amount of available stocks combined with a fall of the stock price signals a strong negative market trend. Therefore, an investor wants to take a long position in the bond. The opposite reasoning holds when an investor wants to sell the bond. I compute monthly volume trading rules for $s = 1,2,3$ and $l = 9,12,15,18,21$.

The last type of technical indicators I study is a momentum based strategy. I construct the trading signal S_t as

$$S_t = \begin{cases} 1 & \text{if } fs_t^{(n)} \geq fs_{t-k/12}^{(n)} \\ 0 & \text{if } fs_t^{(n)} < fs_{t-k/12}^{(n)} \end{cases}. \quad (3)$$

Intuitively, when the current bond price rises relative to its price k periods before, this results

in a positive momentum and relatively high expected excess returns. Therefore, creating a buy signal. I compute monthly momentum rules for $n = 2,3,4,5$ and $k = 3,6,9,12$.

All the earlier constructed trading signals are binary in the sense that one can choose to either buy or sell at the end of period t . However, it is not possible to assess how strong this trading signal is and to what extent it is reasonable for an investor to trade the bond. Therefore, I also construct non-binary trading signals which range from $(-1, 1)$. A signal between -1 (0) and 0 (1) indicates to what extent an investor should sell (buy) the bond. In this way, it is possible to decide for which values of $-1 < a \leq 0$ and $0 \leq b < 1$ it is justified to trade the bond. For all three types of technical indicators, I use the same moving averages and momentum technical indicators as before. However, for the forward spread and volume based moving average I take the difference between the moving average for $j = s$ and $j = t$. For the momentum based strategy, I take the difference between the bond price at the end of period $t - k/12$ and t . Then I normalize the differences such that all values range between $(-1, 1)$.

2.2 Testing the spanning hypothesis

To evaluate the predictive power of the technical indicators, I test the spanning hypothesis by using a regression of the following form

$$y_{t+h} = \beta_1' x_{1t} + \beta_2' x_{2t} + u_{t+h},$$

where y_{t+h} is a future yield or excess bond return, x_{1t} and x_{2t} are vectors consisting of G_1 and G_2 regressors, respectively, and u_{t+h} is a forecast error. The regressors x_{1t} consist of a constant and the information in the yield curve reflected by the level, slope, and curvature. Testing the spanning hypothesis boils down to testing the following null hypothesis

$$H_0 : \beta_2 = 0,$$

which means that the technical indicators, x_{2t} , have no incremental predictive power such that all the relevant information is contained in the yield curve. In practice, including all of the technical predictors in a predictive regression may cause in-sample over-fitting, which most of the times results in very poor out-of-sample forecasting performance. Therefore I use a principal component analysis to still include all the relevant information from the technical indicators while avoiding potential over-fitting. In this analysis, I determine the amount of common factors according to the information criteria developed in Bai and Ng (2002). An advantage of using this criteria relative to the usual AIC and BIC is that the number of common factors can be estimated consistently when both the cross-section dimension and the time dimension of the panel are large.

I analyze the relevance of the technical indicators in two ways: the first is based on the increase in the R^2 of the regression when the technical indicators are included; the second is based on formal statistical tests of the null hypothesis $\beta_2 = 0$. In general, the first three PCs of yields, x_{1t} , are not strictly exogenous because they reflect the information in the current yield curve. Hence the PCs are correlated with u_t . In addition, the predictors are very persistent. Bauer and Hamilton (2018) show that if x_{1t} is not strictly exogenous and if x_{1t} and x_{2t} are highly persistent regressors, then standard tests that do account for heteroskedasticity and autocorrelation (also called HAC standard errors) reject the null hypothesis $\beta_2 = 0$ too often. This bias is also referred to as standard error bias. Furthermore, most predictive regressions in similar settings use monthly data to predict an annual excess return. This causes the use of overlapping returns and results in correlation in the lagged forecast errors u_{t+h} . Therefore, the HAC standard errors are even more biased, and an increase in R^2 is harder to interpret. This suggests that it is of great importance to use methods and statistical tests that are robust to both biases. Therefore, I use two approaches that give more robust small-sample inference. The first procedure is a parametric bootstrap method to test the spanning hypothesis. A second procedure is a statistical test suggested by Ibragimov and Müller (2010) (forth IM-test). Bauer and Hamilton (2018) confirm that both procedures do take into account these significant small-sample problems.

2.2.1 Parametric bootstrap method to test the spanning hypothesis

By bootstrapping under the relevant null hypothesis, it is possible to determine the small-sample size of standard tests and to evaluate their robustness. Moreover, this procedure also allows to test the null of the spanning hypothesis with better precision which should result in more powerful tests and more accurate estimates (Horowitz (2001)). I start the bootstrap procedure by calculating the first three PCs of yields denoted by

$$x_{1t} = (PC1_t, PC2_t, PC3_t)',$$

and by calculating the weighting vector \hat{w}_n for the n -year bond yield:

$$i_{nt} = \hat{w}_n' x_{1t} + \hat{v}_{nt}.$$

Here, $x_{1t} = \hat{W}i_t$, where $i_t = (i_{n_1t}, \dots, i_{n_Kt})'$ is a $K \times 1$ vector of yields at the end of period t , $\hat{W} = (\hat{w}_{n_1}, \dots, \hat{w}_{n_J})'$ is a $3 \times K$ matrix such that the rows are equal to the first three eigenvectors of the variance matrix of i_t , and \hat{v}_{nt} is a fitted error. Furthermore, I normalize the eigenvectors such that $\hat{W}\hat{W}' = I_3$. Afterwards, I estimate a VAR(1) model for x_{1t} by using OLS:

$$x_{1t} = \hat{\phi}_0 + \hat{\phi}_1 x_{1,t-1} + e_{1t} \quad t = 1, \dots, T,$$

from which I generate 5000 bootstrap yield samples. All samples have length T which is equal to the length of the original sample. I start the recursion by drawing from the unconditional distribution for x_{1t} which is estimated from the following VAR(1) model:

$$x_{1\tau}^* = \hat{\phi}_0 + \hat{\phi}_1 x_{1,\tau-1}^* + e_{1\tau}^*.$$

Here, $e_{1\tau}^*$ is the bootstrap residual and the first iteration of every bootstrap sample is equal to the starting value of the observed sample. Next I realize the bootstrap yields using

$$i_{n\tau}^* = \hat{w}'_n x_{1\tau}^* + v_{n\tau}^* \quad \text{for } v_{n\tau}^* \sim N(0, \sigma_v^2),$$

where the standard deviation of the measurement errors, σ_v , is equal to the sample standard deviation of the fitted errors \hat{v}_{nt} . In this way, the parametric bootstrap generates bootstrap yield samples $i_{n\tau}^*$ which are only realized from the three yield curve factors in $x_{1\tau}^*$. On the other hand, the structure of the covariance matrix and its dynamics are comparable to those of the observed bond yield data i_{nt} .

For the technical indicators x_{2t} I use a similar VAR(1) model

$$x_{2t} = \hat{\alpha}_0 + \hat{\alpha}_1 x_{2,t-1} + e_{2t}.$$

I then obtain 5000 bootstrap samples $x_{2\tau}^*$ in the same way as done for $x_{1\tau}^*$.

In order to compute the size and power of the bootstrap test, I use the Monte Carlo simulation developed in Bauer and Hamilton (2018).

2.2.2 IM-test

A second procedure that would give considerably more robust estimates of the small-sample distribution of the test statistics is the IM-test. Ibragimov and Müller (2010) developed a method for robust inference about a scalar coefficient when the data is heterogeneous and correlated in a largely unknown way. The original dataset is partitioned into r subsamples after which the statistic is estimated separately over each subsample. When the estimates are approximately independent and normally distributed across subsamples, then a conventional t -test with r degrees of freedom is used to test the hypothesis $\beta_2 = 0$. According to Müller (2014), the size of the IM-test is exceptional in settings where standard HAC inference would result in poor estimates of the true small-sample variance. In this paper, I only consider 8 and 16 subsamples (as in Müller (2014)).

I use the same Monte Carlo simulation as for the bootstrap procedure to compute the size and power of the IM-test. An advantage of using this test over a standard t -test is that the IM-test has little to no standard error bias. This is because the IM-test results in more precise

estimates of the sampling variability of the test statistic since it splits the sample into multiple subsamples.

2.3 Economic value from an asset allocation perspective

A relatively large increase in R^2 by including the technical indicators may be of little interest for an investor (Thornton and Valente (2012)). On the other hand, utility gain is one of the main measurements used by investors to assess the economic value of any item. Therefore I study the economic value of the bond risk premia forecasts by considering utility gains from an asset allocation perspective. Specifically, I compute the utility gains of a mean-variance investor who has a risk aversion coefficient of three, similar to Goh et al. (2013). This investor optimally allocates a portfolio between a one-year risk-free Treasury bill and an n -year discount bond every month. The allocation decision of the investor is only based on the excess return forecasts obtained by a predictive regression model which includes (a combination of) the PCs of the yield curve, technical indicators and macro-economic factors (forth combined model). At the end of period t , the investor compares this forecast to the historical average forecast and allocates his wealth according to the following weighting scheme

$$\tilde{w}_{t+1}^{(n)} = \frac{1}{\gamma} \frac{\tilde{r}_{t+1}^{(n)}}{\tilde{\sigma}_{n,t+1}^2}. \quad (4)$$

Here, $\tilde{w}_{t+1}^{(n)}$ is the fraction of wealth which he invests in an n -year discount bond during period $t + 1$, γ is the risk aversion coefficient, $\tilde{r}_{t+1}^{(n)}$ is a forecast for the n -year excess bond return, and $\tilde{\sigma}_{n,t+1}^2$ is a forecast for the n -year excess bond return variance. To estimate the variance, the investor bases his decision on a moving window of four years including historical excess bond returns. Afterwards, the average utility of the investor who uses forecasts based on the combined model can be calculated as

$$\hat{v}^{(n)} = \hat{\mu}_n - 0.5\gamma\hat{\sigma}_n^2,$$

where $\hat{\mu}_n$ and $\hat{\sigma}_n^2$ are the sample mean and variance, respectively, of the portfolio constructed based on the weights in (4).

The same procedure can be used for an investor who instead bases his allocation decision on the historical average forecast. At the end of period t , this investor allots his wealth according to a similar weighting scheme

$$\bar{w}_{t+1}^{(n)} = \frac{1}{\gamma} \frac{\bar{r}_{t+1}^{(n)}}{\bar{\sigma}_{n,t+1}^2}. \quad (5)$$

Here, $\bar{w}_{t+1}^{(n)}$ is the fraction of wealth which he invests in a one-year risk-free Treasury bill during period $t + 1$, and $\bar{r}_{t+1}^{(n)}$ is a forecast for the n -year historical average excess return. Next, the

average utility of this investor can be computed as

$$\bar{v}^{(n)} = \bar{\mu}_n - 0.5\gamma\bar{\sigma}_n^2,$$

where $\bar{\mu}_n$ and $\bar{\sigma}_n^2$ are the sample mean and variance, respectively, of the portfolio constructed based on the weights in (5).

The utility gain, $\hat{v}^{(n)} - \bar{v}^{(n)}$, can be explained as the portfolio management fee that the investor is willing to pay to gain access to the combined predictive regression model and the bond risk premia forecasts $\hat{r}_{t+1}^{(n)}$.

3 Data

In order to replicate the results of Cochrane and Piazzesi (2005), I use the original data that is provided on the website of the Federal Reserve Bank of San Francisco¹ which ranges from 1964:01 - 2003:12. Since the original sample period, a substantial amount of new data is available for research. Therefore, I assess the true out-of-sample performance of the earlier mentioned models by using two sample periods. The first is equal to the original sample period in Cochrane and Piazzesi (2005). The later sample period ranges from 1985:01 - 2016:12 which includes some influential events like the financial crisis in 2008.

For the construction of the technical indicators, I use the Fama-Bliss discount bond prices for maturities ranging from one to five years. This dataset is accessible at the Center for Research in Securities Prices (CRSP) and spans the period 1964:01 - 2016:12. Furthermore, I compute monthly yields, forward rates and forward spreads as explained in Section 2. I also use monthly forward spreads to construct the forward spread moving average and momentum-based indicators in (1) and (3), respectively. Lastly, to compute the trading volume-based technical indicators in (2), I use monthly volume data for the S&P 500 index from Yahoo Finance.

Table 1 shows summary statistics for the technical indicator PC factors $\hat{f}_{i,t}^{fs}$, $\hat{f}_{i,t}^{OBV}$ and $\hat{f}_{i,t}^{MOM}$. These PC factors are obtained from 48 forward spread indicators MA^{fs} , 15 volume-based indicators MA^{OBV} and 16 momentum based indicators, respectively, spanning the period 1964:01 - 2016:12. The top panel shows results for the binary trading signal; the bottom panel reports results for the non-binary signal. In the top panel, the first row of Panel $\hat{f}_{i,t}^{fs}$ shows that the first PC factor accounts for 54% of the total variance in the MA^{fs} indicators. The inclusion of the second and third PC factor increases the explained variance to 75%. For $\hat{f}_{i,t}^{OBV}$ and $\hat{f}_{i,t}^{MOM}$, the first PC factor explains 82% and 50% of the total variance in the MA^{OBV} and momentum based indicators, respectively. Including the first three PC factors increases the total

¹The dataset can be found at <https://www.frbsf.org/economic-research/economists/michael-bauer/>

explained variance to 92% and 68%, respectively. For the non-binary trading signal, only two PCs are sufficient to explain most of the variance in the technical indicators, according to the information criterion of Bai and Ng (2002). In addition, these PCs also explain a larger fraction of the total variance relative to the binary trading signal.

Column $AR1_i$ presents the first-order autoregressive coefficients of an AR(1) model for each factor. In the top panel, the autoregressive coefficients for the forward spread technical PC factors $\hat{f}_{i,t}^{fs}$ range from 0.87-0.97. The trading volume technical PC factors $\hat{f}_{i,t}^{OBV}$ have autoregressive coefficients between 0.16 and 0.92. For the momentum based PC factors, the autoregressive coefficients are slightly lower ranging from 0.56 to 0.85, but decline slower in comparison to the volume based PCs. In the lower panel, similar results hold for the non-binary signal with autoregressive coefficients ranging from 0.78 to 0.99 over all technical PC factors. The relatively high persistence of most technical PC factors combined with the fact that the first three PCs of the yield curve are typically not strictly exogenous indicate that it is indeed necessary to use more robust approaches for inference.

Lastly, the time series plots in figures 1 - 3 show how each technical indicator varies over time. Especially during important financial events such as the stock market crash in 1987, the burst of the dot-com bubble in 2000 and the financial crisis in 2008, the forward spread moving average and the bond prices are substantially lower. In addition, the volume-based moving average displays a significant increase during the same period as the financial crisis. This indicates a sudden rise in the amount of available stocks combined with a fall of the stock price which causes a negative stock market trend. In unreported results, I find the same behavior for different lengths j of the forward spread and volume-based moving average.

4 Robust results Cochrane and Piazzesi (2005)

Cochrane and Piazzesi (2005) (forth CP) found that a linear combination of (lagged) forward rates is significant in forecasting bond risk premia for various maturities. CP found that the addition of this variable gives R^2 of 37% (and close to 44% when including lagged forward rates). In particular, CP showed that the level, slope and curvature did not reflect all the information in the yield curve but that higher-order PCs had additional predictive power. Since the first type of technical indicators I use is constructed based on (lagged) forward rates, it is necessary to determine whether the results of Cochrane and Piazzesi (2005) are robust to the earlier mentioned problematic features of the predictive regression model. I do this by testing the spanning hypothesis that only the level, slope and curvature predict excess returns and that the fourth and fifth PC do not add any significant information.

The top panel of table 2 shows the (unadjusted) R^2 for the regression of yearly excess bond returns with three and five PCs as regressors, using the original sample period. The first row replicates the original results; the second row shows results using the bootstrap procedure (with 95% confidence intervals in the third row). For the bootstrap procedure, the inclusion of higher-order PCs does not increase the R^2 as much as in the original results. In addition, the increase in R^2 reported by CP is outside the 95% bootstrap interval. This indicates that the original results are quite unconvincing and are not robust to the small-sample size problems.

I report the same statistics for the later sample period in the lower panel of table 2. In this period, the increase in R^2 by including the fourth and fifth PC is significantly smaller when using the method proposed by CP. The increase in R^2 is also inside the 95%-bootstrap interval for any method, which is in contrast to the original sample period.

In the top panel of table 3, I replicate additional bootstrap estimates of CP. The HAC p -value of 0.000 suggests that $PC4$ is indeed a significant predictor such that the spanning hypothesis would be rejected. The same conclusion can be drawn from the results of the Wald test (also equal to 0.000) under the hypothesis that $PC4$ and $PC5$ add no additional information, as shown in the last column of table 3. Remarkable is that the bootstrap procedure can not reject that $PC4$ is significant. This implies that CP's finding is not caused by the earlier mentioned small-sample size problems. The reason for the significance of $PC4$ under the bootstrap is because the persistence of the higher-order PCs is quite low. Therefore, they are not heavily affected by size distortions². However, the IM-tests do suggest that the fourth and fifth PC are not statistically significant predictors, as it can not reject the null hypothesis that $\beta_2 = 0$. Only $PC1$ and $PC2$ are strongly significant predictors of excess bond returns. Moreover, the size and power of the IM tests are estimated to be closer to nominal size (5% and 95%, respectively) compared to conventional tests. This illustrates that the IM test indeed performs very well in case there is standard-error bias and overlapping return bias and that the results of the IM-test are more convincing than those of the bootstrap procedure.

Similar bootstrap results for the later sample period can be found in the bottom panel of table 3. According to the p -values of the HAC t -test, only $PC1$ and $PC2$ are significant. In particular, the higher-order PCs are not significant for any estimation method (including the method used in CP). This can also be seen when considering the p -values of the Wald test, which are considerably higher for the later sample period. The IM tests imply that only the level is a significant predictor.

Lastly, a substantial amount of data is available since the results of CP in 2005. Therefore

²For example, the first order autocorrelation of $PC4$ and $PC5$ are 0.425 and 0.227, while the twelfth order autocorrelation is equal to 0.062 and -0.135, respectively.

I analyze the true out-of-sample predictive power of the higher-order PCs, which are presented in table 4. The first and second column show the in-sample R^2 for the model using only the first three PCs and for the model including all five PCs, respectively. While the inclusion of the higher-order PCs reduces the in-sample MSE by 11%, the out-of-sample predictive power of this model is worse as the MSE increases by 21%. Although the Diebold-Mariano test does not reject the null hypothesis that both models have equal predictive accuracy, figure 4 shows that only using the first three PCs of the yield curve gives more stable and more precise forecasts of excess bond returns. During most time periods the unrestricted model containing all five PCs estimated expected excess returns that were significantly farther away from the realized excess return than those estimated by the restricted model. This holds especially in high volatile periods such as the financial crisis in 2008. It is remarkable, however, that the unconditional mean (which is estimated over the CP sample period) ended up to be better in predicting future excess returns than both the unrestricted and restricted model.

These results suggest that the forecasting performance of the fourth and fifth PC is sample dependent and that the spanning hypothesis for estimating robust bond risk premia is not rejected.

5 Forecasting excess returns using technical indicators

Before I evaluate the predictive power of the technical indicators, I first compare whether the binary or the non-binary specification of the trading signal results in more significant predictors. Tables 12 and 13 in the Appendix show results for both signals including the three different types of indicators. Considering the p -values of the HAC t -test with standard asymptotic critical values, both specifications show significant technical PCs for the two moving average indicators ($\hat{f}_{1,t}^{fs}$, $\hat{f}_{2,t}^{fs}$, and $\hat{f}_{1,t}^{OBV}$ for the binary signal, and $\hat{f}_{2,t}^{OBV}$ for the non-binary signal). This significance is substantially weaker when employing the bootstrap under the spanning hypothesis, where only one technical PC is still significant ($\hat{f}_{1,t}^{fs}$ and $\hat{f}_{2,t}^{OBV}$, respectively). However, for both specifications the Wald test does not reject the null of the spanning hypothesis suggesting that the technical PCs are not important in predicting bond returns. The p -values for the IM-test are also in line with this conclusion as only the level of the yield curve is significant.

Contrary to the first two indicators, the momentum based strategy is not significant in any of the two specifications. Hence, including only the first two types of indicators may further improve the predictive power of the technical PCs. Table 5 shows results of this binary specification. According to the p -values of the HAC t -test, only the slope, curvature and the technical PCs $\hat{f}_{1,t}^{fs}$ and $\hat{f}_{2,t}^{fs}$ are significant. The Wald statistic also rejects the spanning hypothesis with a p -value

exceeding the critical value for conventional significant levels.

However, the bootstrap procedure shows that none of the technical PCs are robust predictors. The Wald test does not reject the spanning hypothesis as well, which results in the same conclusion. When I consider the results of the IM-tests, only the level and the slope of the yield curve are significant at the five percent level suggesting that just the level and slope are robustly significant predictors.

Regarding the size and power, it is seen that the true size of the standard t -tests range between 8-16% which is substantially higher than the intended five percent. Because all of the individual technical PC factors have a non-standard small-sample distribution, the size distortion for the Wald test is even larger with an actual size of 36%. However, both the bootstrap and the IM-test are close to nominal size. Especially the IM-test has relative good power.

When one compares these results with the non-binary specification in table 6, one can notice that under the bootstrap the second PC of the volume based indicator $\hat{f}_{2,t}^{OBV}$ is still significant (even at the one percent level). However, this evidence is still not strong enough to reject the spanning hypothesis according to the Wald test. The IM-tests for this specification show slightly different results compared to the binary case. In this case, only the level is a robust predictor while both the level and the slope were robust predictors previously. The size and power show similar patterns for both specifications.

In the later sample period, the level is a significant predictor under the binary specification with a HAC p -value of 0.047. However in the non-binary case, it is remarkable that the level, slope and curvature are not significant (not even at the ten percent level). This result heavily contradicts the literature and should raise concerns on whether the non-binary trading signal is an appropriate specification in this sample period.

Lastly I study the increase in R^2 when including the technical PCs. Table 7 (binary signal) and table 8 (non-binary signal) show that the inclusion of the technical PCs indeed increases the R^2 of the regression. When I employ the bootstrap procedure, the increase in R^2 is smaller for both types of signals in the later sample but higher for the non-binary signal in the original sample. Recall that the bootstrap is constructed such that the technical PCs add no additional information in predicting future excess returns. Hence, the latter observation is quite outstanding and most likely suggests that a non-binary trading signal does not entirely summarize the information in the technical indicators. Therefore, I only consider the binary trading signal including the forward spread and volume based moving average when I analyze the forecasting performance of technical indicators combined with macro-economic factors in Section 6.

Overall, these results do not indicate sufficient evidence against the spanning hypothesis.

Although some technical PCs are significant using the bootstrap, the results of the Wald test and IM-test consistently show that the first two PCs of yields are the only robust predictors of excess bond returns.

6 Forecasting excess returns using technical indicators and macroeconomic factors

The predictability of technical indicators can possibly be improved upon after controlling for macro-economic factors. Cochrane and Piazzesi (2005) provide strong empirical evidence that the current term structure does not span all information needed to forecast future excess returns and that the fourth and fifth PC contain additional predictive information. Therefore, I use the model specification as chosen in Section 5 and combine this with the two higher-order PCs of the yield curve to evaluate whether this model improves the forecasting performance of the technical indicators.

The top panel of table 9 shows the (unadjusted) R^2 for this regression, using the original sample period. The increase in R^2 is considerably higher in the original data (barely inside the 95% bootstrap interval) relative to the bootstrap procedure. In addition, the inclusion of higher-order PCs increases the R^2 to 0.42 in the original data compared to a R^2 of 0.34 for the model with only the technical indicators as additional predictors (table 7). Therefore, the inclusion of higher-order PCs increases the R^2 with 0.08. Under the bootstrap, the increase in R^2 is significantly lower with an increase of 0.01. This suggests that controlling for macro-economic factors, in the form of the fourth and fifth PC of the yield curve, does not improve the forecasting performance of technical indicators in terms of increase in R^2 .

In the later sample period, the increase in R^2 is even larger in the original data and is outside the range of the 95% bootstrap interval. Surprisingly, the addition of $PC4$ and $PC5$ does not increase the R^2 in this sample period as the R^2 in table 7 and 9 are the same (under both the original data and the bootstrap). Therefore, higher-order PCs most likely do not add any additional predictive information in forecasting excess bond returns.

Additional bootstrap results for the original sample period are reported in the top panel of table 10. The HAC p -values of the t -test indicate that the higher-order PCs as well as the first two PCs of the forward spread moving average are significant predictors, and that the spanning hypothesis would be rejected. When I compare these results with the p -values under the bootstrap procedure, quite similar results can be found as the robust results of Cochrane and Piazzesi (2005) in Section 4. For instance, the higher-order PCs and the first PC of the forward

spread moving average, $\hat{f}_{1,t}^{fs}$, are still significant, and the spanning hypothesis is also rejected. The significance of the higher-order PCs is most likely due to the same reasons as discussed in Section 4. While the Wald statistic rejects the spanning hypothesis under the bootstrap as well, this evidence is weaker relative to the Wald statistic under the HAC t -test.

However, the p -values of the IM-test indicate that only $PC1$ and $PC2$ are significant predictors. In particular, the p -values for the predictors that were significant under the bootstrap ($PC4$, $PC5$, and $\hat{f}_{1,t}^{fs}$) are far greater than the conventional significance levels. Therefore, I can conclude that only the first two PCs of the yield curve are robust predictors.

Table 10 also reports the size and power of the different tests. The results reveal that the standard HAC t -test has serious size distortions (ranging from 9% to 17%) which causes the size of the Wald statistic to be even greater. Meanwhile, the bootstrap procedure and both IM-tests show size levels that are much closer to the correct size of five percent. Moreover, the IM-tests perform very well in terms of power. This gives me added reason to pay attention to the results of the IM-tests relative to those of the bootstrap procedure.

The lower panel of table 10 shows results for the later sample period. In this period, the higher-order and technical PCs are considerably less significant. Under the bootstrap, none of the additional predictors are significant such that the spanning hypothesis would not be rejected. The p -values of the IM-test show similar results as for the original sample period. The level and slope are the only robustly significant predictors.

From these results, I can conclude that including macro-economic factors, in the form of higher-order PCs of the yield curve, does not improve the forecasting performance of technical indicators. Furthermore, the significance of the technical indicators seems to be considerably weaker in the later sample period.

7 Economic value of bond risk premia forecasts

In this section, I evaluate the economic value of the bond risk premia forecasts for a mean-variance investor who has a risk aversion coefficient of three. Every month, the investor optimally manages a portfolio containing an n -year discount bond and a one-year Treasury bill. Table 11 displays the average utility gains, annualized in percentage points, for portfolios constructed with various combinations of predictors. The average utility gain can be interpreted as the portfolio management fee that the investor is willing to pay to gain access to the combined predictive regression model and the bond risk premia forecasts. The results show that a forecasting model based on only the technical PCs, $\hat{f}_{1,t}^{fs} - \hat{f}_{3,t}^{OBV}$, is of considerable economic value with a maximum utility gain of 2.51% per annum. In the later sample period, the performance of this portfolio

becomes worse with the utility gain ranging between 0.48% and 2.12%.

However, a forecasting model based on the level, slope and curvature is able to generate more economic value with a maximum utility gain of 2.78%. The inclusion of higher-order PCs ($PC1 - PC5$) slightly improves the forecasting model in the original sample period with a maximum utility gain of 2.90%. In the later sample period, this portfolio performs worse relative to a portfolio based on only the first three PCs of yields as the maximum utility gain is only 2.00% (compared to 2.45% for the latter portfolio). This is in line with my main findings in Section 4 (table 3) where I found that $PC4$ was a significant predictor in the original sample period but both higher-order PCs were insignificant in the later sample period. Therefore, higher-order PCs generate little economic value in the later sample period as expected.

Next, I consider the economic gains of forecasting models which combine the information in the technical indicators and the macro-economic variables (the last two columns of table 11). For all short-term discount bonds, using the 5 PCs of yields and technical PCs as predictors results in higher average utility gains in both sample periods than by not including higher-order PCs. This result is consistent with my conclusion in Section 6 (table 10) where I found that several technical PCs are still significant after controlling for macro-economic factors in the form of higher-order PCs. Therefore, I expected that a forecasting model based on technical indicators and macro-economic variables would still generate positive economic value.

It is worth emphasizing that the average utility gains in the later sample period are considerably lower relative to the original sample period. This is most likely due to the volatile bond prices during the financial crisis in 2008. Therefore, bond risk premia are harder to forecast which causes a decrease in utility gains for all different forecasting models.

Overall, a portfolio based on the first five PCs of the yield curve and the technical indicators generates the most utility gains. An investor would be willing to pay an annual management fee of close to 3.0% to gain access to the 5-year excess bond return forecasts generated by this model.

8 Conclusion

In this paper, I investigate the importance of technical indicators in forecasting robust bond risk premia. I find that technical indicators are significant but not robust predictors of excess bond returns. Furthermore, I study whether including macro-economic factors, in the form of higher-order PCs, improves the predictive power. My results show that higher-order PCs do not increase the forecasting performance of technical indicators. A common finding in all my results is that the level and slope of the yield curve are the only robust predictors of bond risk premia.

From an asset allocation perspective, I find that forecasts obtained from the first five PCs of yields and the technical indicators generate the most utility gains.

A natural extension of this research is to study the forecasting power of different technical indicators in combination with other macro-economic factors. For example, in addition to the trend-following moving averages constructed in this paper, Wong et al. (2003) also focus on the most regularly used counter-trend indicator, the so-called Relative Strength Index. Another idea is to use a different set-up for the non-binary trading signals. In this paper, the non-binary signal is defined based on the difference between two technical indicators (of the same type). It is also possible to take into account the volatility during unstable periods by scaling the moving averages by their (downside) volatility. The higher the (downside) volatility, the less reliable the trading signal. Incorporating these fundamental and technical indicators in term-structure models of bond pricing is left for future research.

9 References

- Bai, Jushan, and Serena Ng. "Determining the number of factors in approximate factor models." *Econometrica* 70.1 (2002): 191-221.
- Bauer, Michael D., and James D. Hamilton. "Robust bond risk premia." *The Review of Financial Studies* 31.2 (2018): 399-448.
- Campbell, John Y., and Tuomo Vuolteenaho. "Inflation illusion and stock prices." *American Economic Review* 94.2 (2004): 19-23.
- Cochrane, John H., and Monika Piazzesi. "Bond risk premia." *American Economic Review* 95.1 (2005): 138-160.
- Diebold, Francis X., and Robert S. Mariano. "Comparing predictive accuracy." *Journal of Business & economic statistics* 20.1 (2002): 134-144.
- Fama, Eugene F., and Kenneth R. French. "Business conditions and expected returns on stocks and bonds." *Journal of financial economics* 25.1 (1989): 23-49.
- Goh, Jeremy, et al. "Forecasting government bond risk premia using technical indicators." (2013).
- Horowitz, Joel L. "The bootstrap." *Handbook of econometrics*. Vol. 5. Elsevier, 2001. 3159-3228.
- Ibragimov, Rustam, and Ulrich K. Müller. "t-Statistic based correlation and heterogeneity robust inference." *Journal of Business & Economic Statistics* 28.4 (2010): 453-468.
- Litterman, Robert, and Jose Scheinkman. "Common factors affecting bond returns." *Journal of fixed income* 1.1 (1991): 54-61.

Ludvigson, Sydney C., and Serena Ng. "Macro factors in bond risk premia." *The Review of Financial Studies* 22.12 (2009): 5027-5067.

Müller, Ulrich K. "HAC corrections for strongly autocorrelated time series." *Journal of Business & Economic Statistics* 32.3 (2014): 311-322.

Neely, C. J., et al. Forecasting the equity risk premium: The role of technical indicators. Federal Reserve Bank of St. Louis working paper, 2012.

Thornton, Daniel L., and Giorgio Valente. "Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective." *The Review of Financial Studies* 25.10 (2012): 3141-3168.

Wong, Wing-Keung, Meher Manzur, and Boon-Kiat Chew. "How rewarding is technical analysis? Evidence from Singapore stock market." *Applied Financial Economics* 13.7 (2003): 543-551.

Table 1: Summary statistics technical indicator PC factors

i	$\hat{f}_{i,t}^{fs}$	$\hat{f}_{i,t}^{OBV}$		$\hat{f}_{i,t}^{MOM}$		
	$\sum_{j=1}^i \lambda_j / \sum_{j=1}^N \lambda_j$	$AR1_i$	$\sum_{j=1}^i \lambda_j / \sum_{j=1}^N \lambda_j$	$AR1_i$	$\sum_{j=1}^i \lambda_j / \sum_{j=1}^N \lambda_j$	$AR1_i$
<i>Binary trading signal</i>						
1	0.54	0.97	0.82	0.92	0.50	0.85
2	0.68	0.89	0.89	0.63	0.61	0.61
3	0.75	0.87	0.92	0.16	0.68	0.56
<i>Non-binary trading signal</i>						
1	0.92	0.99	0.94	0.96	0.83	0.91
2	0.97	0.96	0.98	0.89	0.93	0.78

Table 1 shows summary statistics for the first three technical indicator principal component (PC) factors $\hat{f}_{i,t}^{fs}$, $\hat{f}_{i,t}^{OBV}$ and $\hat{f}_{i,t}^{MOM}$. These PC factors are obtained from 48 forward spread indicators MA^{fs} , 15 volume-based indicators MA^{OBV} and 16 momentum based technical indicators, respectively, spanning the period 1964:01 - 2016:12. The top panel shows results for the binary trading signal; the bottom panel reports results for the non-binary signal. The second, fourth and sixth column show the fraction of the total variance in the technical indicators explained by PC factor i , which is calculated by dividing the sum of the first i largest eigenvalues λ_i by the sum of all N eigenvalues. Column $AR1_i$ displays the first-order autocorrelation for PC factor i estimated from an AR(1) model. The amount of factors is decided by the information criterion developed by Bai and Ng (2002).

Figure 1: Time series of forward spread moving average rule MA^{fs}

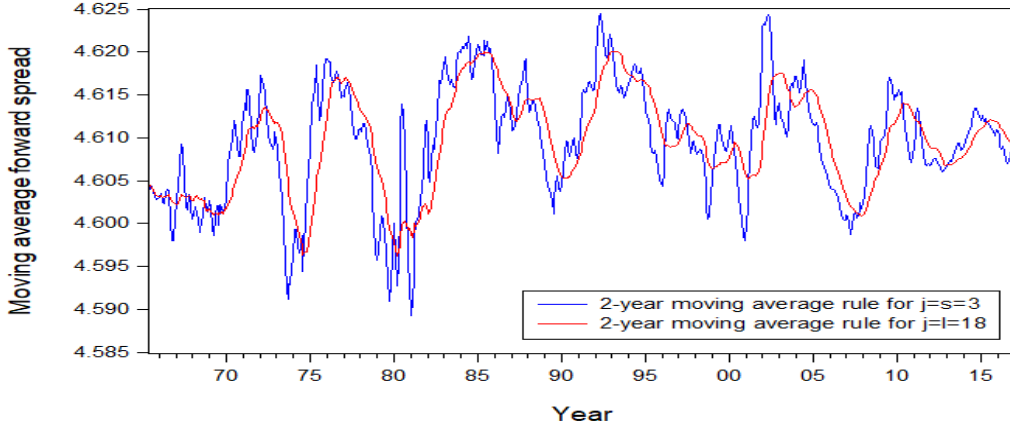


Figure 1 shows a time series of the forward spread moving average rule MA^{fs} , for $j = 3$ and 18. The time series ranges from 1964-01 to 2016-12, containing several significant financial events such as the stock market crash in 1987, the burst of the dot-com bubble in 2000, and the financial crisis in 2008.

Figure 2: Time series of volume-based trading rule MA^{OBV}

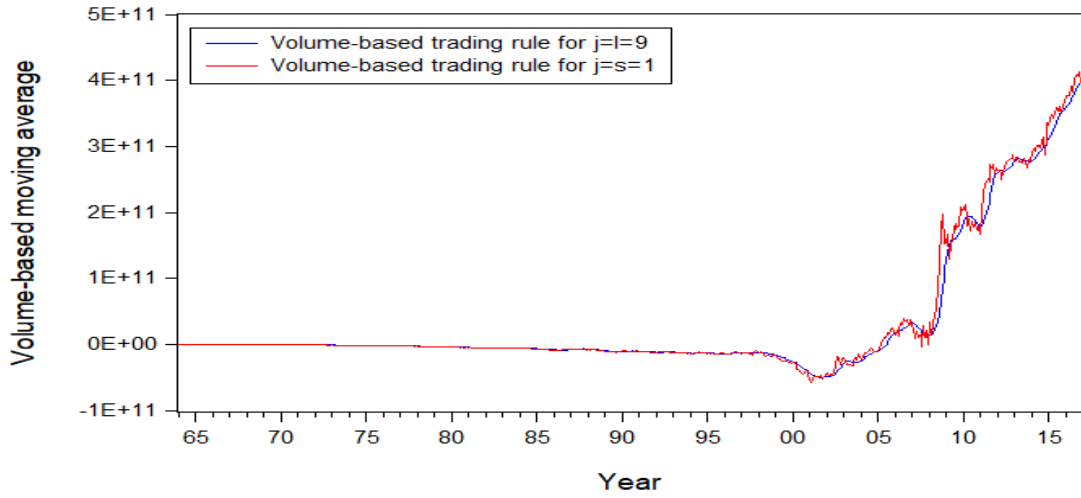


Figure 2 shows a time series of the volume-based moving average rule MA^{OBV} , for $j = 1$ and 9 . The time series ranges from 1964-01 to 2016-12, containing several significant financial events such as the stock market crash in 1987, the burst of the dot-com bubble in 2000, and the financial crisis in 2008.

Figure 3: Time series of n -year bond prices

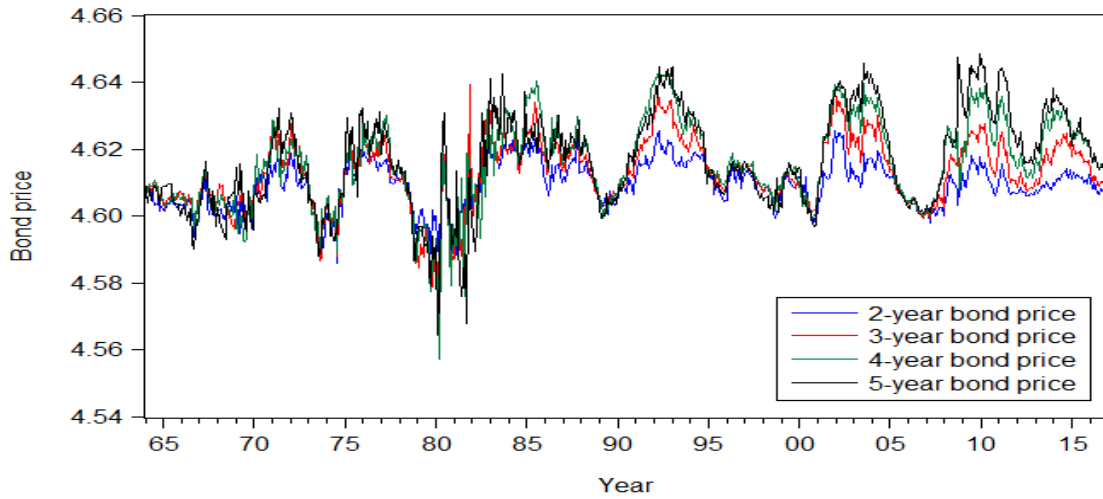


Figure 3 shows a time series of the n -year bond prices. The time series ranges from 1964-01 to 2016-12, containing several significant financial events such as the stock market crash in 1987, the burst of the dot-com bubble in 2000, and the financial crisis in 2008.

Table 2: In-sample adjusted R^2 values of the results in Cochrane-Piazzesi (2005)

	R_1^2	R_2^2	$R_2^2 - R_1^2$
<i>Original sample: 1964-2003</i>			
Data	0.26	0.35	0.09
Bootstrap	0.21	0.22	0.01
95% confidence interval	(0.05, 0.39)	(0.06, 0.40)	(0.00, 0.02)
<i>Later sample: 1985-2016</i>			
Data	0.15	0.18	0.03
Bootstrap	0.30	0.31	0.01
95% confidence interval	(0.10, 0.51)	(0.11, 0.52)	(0.00, 0.05)

Table 2 shows the (unadjusted) R^2 for the regression of yearly excess bond returns (averaged across two through five years) with the level, slope and curvature (R_1^2) and including all five PCs of yields (R_2^2). The third column also displays the difference in R^2 . The upper panel reports results for the original sample period as in Cochrane-Piazzesi (2005); the bottom panel is for the sample period 1985-2016. The first row of both panels shows the R^2 value using the same data as in the original paper. The second row shows the R^2 value according to the bootstrap procedure, and the third row the 95% confidence interval. For the bootstrap method, the null hypothesis is that the higher-order PCs of the yield curve have no additional predictive power.

Table 3: Robust estimation of results in Cochrane-Piazzesi (2005)

	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>	<i>PC4</i>	<i>PC5</i>	Wald
<i>Original sample: 1964-2003</i>						
Coefficient	0.127	2.740	-6.307	-16.128	-2.038	
HAC <i>p</i> -value	0.085	0.000	0.003	0.000	0.455	0.000
Bootstrap <i>p</i> -value				0.000	0.501	0.000
IM <i>q</i> = 8	0.006	0.005	0.717	0.484	0.626	
IM <i>q</i> = 16	0.000	0.040	0.087	0.600	0.477	
<i>Estimated size of tests</i>						
HAC				0.091	0.077	0.112
Bootstrap				0.060	0.040	0.050
IM <i>q</i> = 8				0.053	0.052	
IM <i>q</i> = 16				0.056	0.052	
<i>Estimated power of tests</i>						
HAC				0.997	0.148	0.994
Bootstrap				0.995	0.113	0.988
IM <i>q</i> = 8				0.994	0.589	
IM <i>q</i> = 16				0.998	0.596	
<i>Later sample: 1985-2016</i>						
Coefficient	0.106	1.589	3.157	-9.585	-9.360	
HAC <i>p</i> -value	0.048	0.025	0.343	0.145	0.207	0.124
Bootstrap <i>p</i> -value				0.239	0.295	0.264
IM <i>q</i> = 8	0.040	0.506	0.169	0.211	0.232	
IM <i>q</i> = 16	0.001	0.049	0.671	0.352	0.793	

Table 3 shows bootstrap results of forecasting yearly excess bond returns (averaged across two through five years) with the level, slope and curvature of the yield curve. The results in the upper panel are identical to those of Cochrane and Piazzesi (2005). *p*-values of the HAC *t*-test are obtained using Newey-West standard errors with 18 lags. The column "Wald" shows *p*-values for the χ^2 test that the fourth and fifth PC have no predictive power. I also display results for the IM-test, designed by Ibragimov and Müller (2010), for $r = 8$ and 16 subsamples. Estimates of the size and power are shown in panel *Size* and *Power*, respectively. These estimates are obtained from the Monte Carlo simulation described in Bauer and Hamilton (2018). *p*-values below 5% are highlighted in bold.

Table 4: Out-of-sample forecasting performance of results in Cochrane-Piazzesi

In-sample		Out-of-sample		
R_1^2	R_2^2	MSE-ratio	MSE-ratio	DM p -value
0.267	0.344	0.891	1.213	0.103

Table 4 reports the in-sample and out-of-sample forecasting performance of the model containing the level, slope and curvature as regressors (R_1^2), as well as the model including all five PCs of yields (R_2^2) as regressors. The in-sample period ranges from 1964 to 2002, which is equal to the last observation used by Cochrane and Piazzesi (2005). The out-of-sample period is from 2003 to 2016. The columns show in-sample R^2 for both models, the in-sample mean-squared-errors (MSE) ratio calculated as $\frac{MSE_1}{MSE_2}$, the out-of-sample MSE ratio, and the p -value of the Diebold-Mariano (DM) test of equal forecast accuracy.

Figure 4: Cochrane-Piazzesi: out-of-sample forecasts

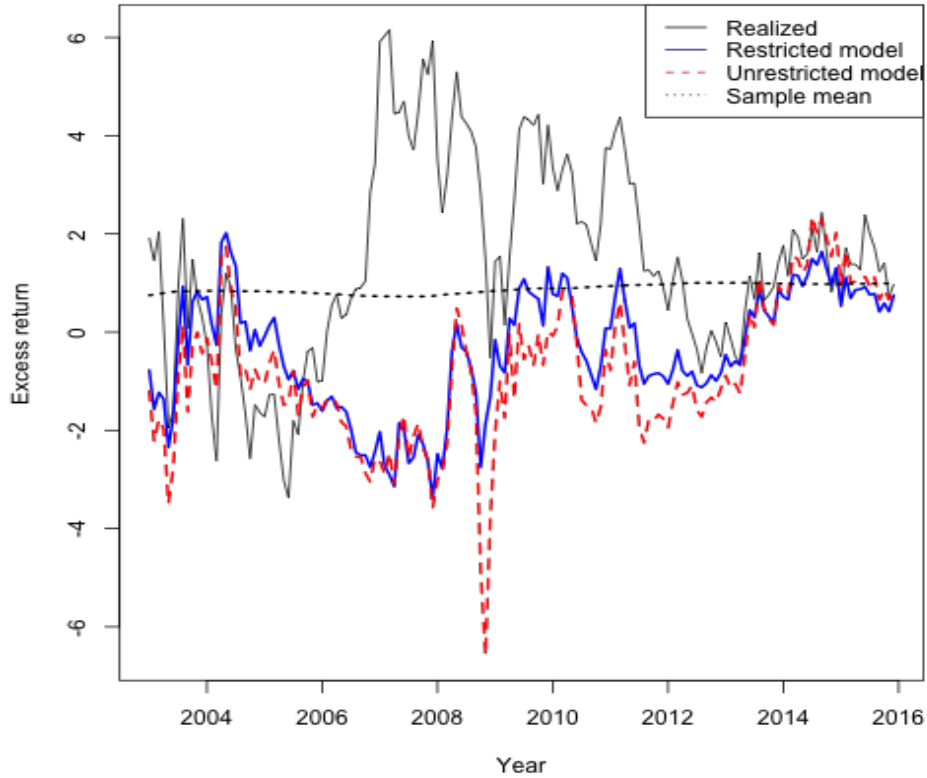


Figure 4 shows realized and predicted excess bond returns from the model with three PCs (restricted model) and the model with all five PCs (unrestricted model). The in-sample period ranges from 1964 - 2002, and the out-of-sample period ranges from 2003 - 2016. The black dashed line shows the in-sample mean excess return.

Table 5: Statistical inference of technical indicators with binary trading signals

	$PC1$	$PC2$	$PC3$	$\hat{f}_{1,t}^{fs}$	$\hat{f}_{2,t}^{fs}$	$\hat{f}_{3,t}^{fs}$	$\hat{f}_{1,t}^{OBV}$	$\hat{f}_{2,t}^{OBV}$	$\hat{f}_{3,t}^{OBV}$	Wald
<i>Original sample: 1964-2003</i>										
Coefficient	0.134	1.227	-5.018	0.850	-0.853	0.533	0.597	-1.348	-0.792	
HAC p -value	0.224	0.041	0.007	0.010	0.027	0.397	0.052	0.155	0.221	0.001
Bootstrap p -value				0.065	0.119	0.509	0.124	0.243	0.284	0.103
IM $q = 8$	0.002	0.010	0.857	0.788	0.293	0.418	0.610	0.892	0.377	
IM $q = 16$	0.000	0.000	0.139	0.263	0.347	0.351	0.118	0.753	0.621	
<i>Estimated size of tests</i>										
HAC				0.149	0.164	0.125	0.120	0.105	0.084	0.362
Bootstrap				0.060	0.061	0.060	0.046	0.050	0.049	0.054
IM $q = 8$				0.079	0.068	0.077	0.068	0.067	0.061	
IM $q = 16$				0.074	0.066	0.074	0.072	0.068	0.065	
<i>Estimated power of tests</i>										
HAC				0.661	0.366	0.201	0.638	0.378	0.235	0.914
Bootstrap				0.466	0.176	0.095	0.512	0.267	0.178	0.598
IM $q = 8$				0.723	0.691	0.582	0.788	0.710	0.694	
IM $q = 16$				0.712	0.654	0.613	0.756	0.647	0.593	
<i>Later sample: 1985-2016</i>										
Coefficient	0.310	-0.570	0.326	1.732	-0.887	0.859	0.189	-0.315	-0.260	
HAC p -value	0.047	0.350	0.921	0.000	0.165	0.213	0.646	0.609	0.576	0.000
Bootstrap p -value				0.009	0.287	0.383	0.745	0.691	0.635	0.166
IM $q = 8$	0.002	0.033	0.429	0.188	0.557	0.308	0.069	0.204	0.772	
IM $q = 16$	0.024	0.045	0.883	0.585	0.304	0.475	0.359	0.569	0.765	

Table 5 shows bootstrap results of forecasting yearly excess bond returns (averaged across two through five years) with the level, slope and curvature of the yield curve, the first three technical PC factors of the first two technical indicators $\hat{f}_{i,t}^{fs}$ and $\hat{f}_{i,t}^{OBV}$, and a binary trading signal. p -values of the HAC t -test are obtained using Newey-West standard errors with 18 lags. The column "Wald" shows p -values for the χ^2 test that all technical PC factors have no predictive power. I also display results for the IM-test, designed by Ibragimov and Müller (2010), for $r = 8$ and 16 subsamples. Estimates of the size and power are shown in panel *Size* and *Power*, respectively. These estimates are obtained from the Monte Carlo simulation described in Bauer and Hamilton (2018). p -values below 5% are highlighted in bold.

Table 6: Statistical inference of technical indicators with non-binary trading signals

	$PC1$	$PC2$	$PC3$	$\hat{f}_{1,t}^{fs}$	$\hat{f}_{2,t}^{fs}$	$\hat{f}_{1,t}^{OBV}$	$\hat{f}_{2,t}^{OBV}$	Wald
<i>Original sample: 1964-2003</i>								
Coefficient	0.177	2.157	-5.505	0.224	0.158	3.585	-46.965	
HAC p -value	0.127	0.000	0.002	0.189	0.723	0.379	0.000	0.003
Bootstrap p -value				0.346	0.768	0.504	0.008	0.137
IM $q = 8$	0.001	0.054	0.574	0.262	0.659	0.383	0.486	
IM $q = 16$	0.000	0.012	0.105	0.688	0.662	0.132	0.216	
<i>Estimated size of tests</i>								
HAC				0.167	0.144	0.162	0.130	0.346
Bootstrap				0.047	0.044	0.051	0.056	0.050
IM $q = 8$				0.081	0.076	0.073	0.065	
IM $q = 16$				0.067	0.072	0.065	0.062	
<i>Estimated power of tests</i>								
HAC				0.220	0.138	0.234	0.720	0.759
Bootstrap				0.066	0.052	0.098	0.554	0.288
IM $q = 8$				0.612	0.693	0.629	0.854	
IM $q = 16$				0.591	0.645	0.633	0.901	
<i>Later sample: 1985-2016</i>								
Coefficient	0.159	0.264	1.488	107.846	86.307	-0.097	0.417	
HAC p -value	0.130	0.758	0.435	0.002	0.349	0.891	0.853	0.001
Bootstrap p -value				0.038	0.480	0.921	0.882	0.125
IM $q = 8$	0.001	0.102	0.239	0.010	0.169	0.048	0.739	
IM $q = 16$	0.001	0.213	0.100	0.550	0.116	0.561	0.870	

Table 6 shows bootstrap results of forecasting yearly excess bond returns (averaged across two through five years) with the level, slope and curvature of the yield curve, the first two technical PC factors of the first two technical indicators $\hat{f}_{i,t}^{fs}$ and $\hat{f}_{i,t}^{OBV}$, and a non-binary trading signal. p -values of the HAC t -test are obtained using Newey-West standard errors with 18 lags. The column "Wald" shows p -values for the χ^2 test that all technical PC factors have no predictive power. I also display results for the IM-test, designed by Ibragimov and Müller (2010), for $r = 8$ and 16 subsamples. Estimates of the size and power are shown in panel *Size* and *Power*, respectively. These estimates are obtained from the Monte Carlo simulation described in Bauer and Hamilton (2018). p -values below 5% are highlighted in bold.

Table 7: In-sample adjusted R^2 values of technical indicators using binary trading signals

	R_1^2	R_2^2	$R_2^2 - R_1^2$
<i>Original sample: 1964-2003</i>			
Data	0.28	0.34	0.06
Bootstrap	0.23	0.28	0.05
95% confidence interval	(0.06, 0.42)	(0.11, 0.47)	(-0.00, 0.13)
<i>Later sample: 1985-2016</i>			
Data	0.19	0.41	0.22
Bootstrap	0.32	0.39	0.07
95% confidence interval	(0.11, 0.55)	(0.17, 0.61)	(0.00, 0.19)

Table 7 shows the (unadjusted) R^2 for the regression of yearly excess bond returns (averaged across two through five years) using a binary trading signal, including the level, slope and curvature (R_1^2) and the first two PC factors of the first two technical indicators (R_2^2). The third column also displays the difference in R^2 . The upper panel reports results for the original sample period; the bottom panel is for the sample period 1985-2016. The first row of both panels shows the R^2 value using the same data as in the original paper. The second row shows the R^2 value according to the bootstrap procedure, and the third row the 95% confidence interval. For the bootstrap method, the null hypothesis is that all technical PC factors have no additional predictive power.

Table 8: In-sample adjusted R^2 values of technical indicators using non-binary trading signals

	R_1^2	R_2^2	$R_2^2 - R_1^2$
<i>Original sample: 1964-2003</i>			
Data	0.28	0.30	0.02
Bootstrap	0.23	0.29	0.05
95% confidence interval	(0.06, 0.42)	(0.11, 0.47)	(0.00, 0.15)
<i>Later sample: 1985-2016</i>			
Data	0.17	0.30	0.14
Bootstrap	0.28	0.34	0.06
95% confidence interval	(0.08, 0.49)	(0.14, 0.55)	(0.00, 0.18)

Table 7 shows the (unadjusted) R^2 for the regression of yearly excess bond returns (averaged across two through five years) using a non-binary trading signal, including the level, slope and curvature (R_1^2) and the first two PC factors of the first two technical indicators (R_2^2). The third column also displays the difference in R^2 . The upper panel reports results for the original sample period; the bottom panel is for the sample period 1985-2016. The first row of both panels shows the R^2 value using the same data as in the original paper. The second row shows the R^2 value according to the bootstrap procedure, and the third row the 95% confidence interval. For the bootstrap method, the null hypothesis is that all technical PC factors have no additional predictive power.

Table 9: In-sample adjusted R^2 values of technical indicators and macro-economic factors

	R_1^2	R_2^2	$R_2^2 - R_1^2$
<i>Original sample: 1964-2003</i>			
Data	0.28	0.42	0.14
Bootstrap	0.23	0.29	0.06
95% confidence interval	(0.06, 0.42)	(0.12, 0.48)	(0.00, 0.15)
<i>Later sample: 1985-2016</i>			
Data	0.19	0.41	0.22
Bootstrap	0.32	0.39	0.07
95% confidence interval	(0.11, 0.55)	(0.18, 0.62)	(0.01, 0.20)

Table 9 shows the (unadjusted) R^2 for the regression of yearly excess bond returns (averaged across two through five years) using a binary trading signal, including the level, slope and curvature (R_1^2), the fourth and fifth PC of the yield curve along with the first three PC factors of the first two technical indicators (R_2^2). The third column also displays the difference in R^2 . The top panel reports results for the original sample period; the bottom panel is for the sample period 1985-2016. The first row of both panels shows the R^2 value using the same data as in the original paper. The second row shows the R^2 value according to the bootstrap procedure, and the third row the 95% confidence interval. For the bootstrap method, the null hypothesis is that all technical PC factors have no additional predictive power.

Table 10: Statistical inference of technical indicators and macro-economic factors

	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>	<i>PC4</i>	<i>PC5</i>	$\hat{f}_{1,t}^{fs}$	$\hat{f}_{2,t}^{fs}$	$\hat{f}_{3,t}^{fs}$	$\hat{f}_{1,t}^{OBV}$	$\hat{f}_{2,t}^{OBV}$	$\hat{f}_{3,t}^{OBV}$	Wald
<i>Original sample: 1964-2003</i>												
Coefficient	0.136	1.080	-4.797	7.117	11.577	0.982	-0.924	0.408	0.620	-0.977	-0.656	
HAC <i>p</i> -value	0.147	0.053	0.005	0.035	0.000	0.001	0.015	0.490	0.059	0.228	0.346	0.000
Bootstrap <i>p</i> -value				0.087	0.000	0.024	0.094	0.588	0.142	0.313	0.422	0.003
IM <i>q</i> = 8	0.008	0.048	0.492	0.242	0.455	0.727	0.759	0.101	0.568	0.689	0.999	
IM <i>q</i> = 16	0.000	0.002	0.360	0.522	0.089	0.871	0.131	0.222	0.319	0.066	0.593	
<i>Estimated size of tests</i>												
HAC				0.111	0.132	0.157	0.173	0.132	0.129	0.104	0.093	0.472
Bootstrap				0.049	0.056	0.059	0.063	0.056	0.050	0.057	0.050	0.055
IM <i>q</i> = 8				0.060	0.062	0.080	0.065	0.073	0.067	0.063	0.055	
IM <i>q</i> = 16				0.056	0.059	0.076	0.063	0.069	0.066	0.059	0.059	
<i>Estimated power of tests</i>												
HAC				0.686	0.790	0.764	0.410	0.172	0.541	0.262	0.191	0.990
Bootstrap				0.548	0.644	0.583	0.204	0.079	0.374	0.174	0.133	0.840
IM <i>q</i> = 8				0.743	0.879	0.864	0.712	0.649	0.794	0.642	0.556	
IM <i>q</i> = 16				0.754	0.882	0.802	0.748	0.548	0.772	0.613	0.564	
<i>Later sample: 1985-2016</i>												
Coefficient	0.156	0.359	2.033	-2.232	2.386	1.124	-0.627	0.323	0.026	0.017	-0.050	
HAC <i>p</i> -value	0.042	0.011	0.331	0.632	0.733	0.007	0.216	0.582	0.950	0.974	0.887	0.002
Bootstrap <i>p</i> -value				2.765	2.517	2.887	2.582	2.799	2.758	2.524	2.326	44.382
IM <i>q</i> = 8	0.001	0.013	0.729	0.289	0.324	0.897	0.929	0.303	0.894	0.104	0.592	
IM <i>q</i> = 16	0.000	0.051	0.290	0.216	0.386	0.329	0.293	0.342	0.342	0.787	0.060	

Table 10 shows bootstrap results of forecasting yearly excess bond returns (averaged across two through five years) with the level, slope and curvature of the yield curve, and the fourth and fifth PC as macro-economic factors. The predictive regression also includes the first three technical PC factors of the first two technical indicators $\hat{f}_{i,t}^{fs}$ and $\hat{f}_{i,t}^{OBV}$, and a binary trading signal. *p*-values of the HAC *t*-test are obtained using Newey-West standard errors with 18 lags. The column "Wald" shows *p*-values for the χ^2 test that all technical PC factors have no predictive power. I also display results for the IM-test, designed by Ibragimov and Müller (2010), for $r = 8$ and 16 subsamples. Estimates of the size and power are shown in panel *Size* and *Power*, respectively. These estimates are obtained from the Monte Carlo simulation described in Bauer and Hamilton (2018). *p*-values below 5% are highlighted in bold.

Table 11: Asset Allocation Results

n	$\hat{f}_{1,t}^{fs} - \hat{f}_{3,t}^{OBV}$	$PC1 - PC3$	$PC1 - PC5$	$(PC1 - PC3) + (\hat{f}_{1,t}^{fs} - \hat{f}_{3,t}^{OBV})$	$(PC1 - PC5) + (\hat{f}_{1,t}^{fs} - \hat{f}_{3,t}^{OBV})$
<i>Original sample: 1964-2003</i>					
2	0.65	0.87	1.13	1.01	1.29
3	1.49	1.66	1.82	1.84	2.01
4	1.88	2.01	2.14	2.29	2.33
5	2.51	2.78	2.90	3.04	3.08
<i>Later sample: 1985-2016</i>					
2	0.48	0.67	0.52	0.83	1.04
3	0.93	1.45	1.20	1.73	1.92
4	1.63	1.89	1.45	2.11	2.24
5	2.12	2.45	2.00	2.67	2.81

Table 11 shows the average utility gains, annualized in percentage points, for a mean-variance investor who has a risk aversion coefficient of three. Every month, the investor optimally manages a portfolio containing an n -year discount bond and a one-year Treasury bill. The utility gain can be interpreted as the portfolio management fee that the investor is willing to pay to gain access to the combined predictive regression model and the bond risk premia forecasts $\tilde{r}_{t+1}^{(n)}$. The PC factors $\hat{f}_{1,t}^{fs} - \hat{f}_{3,t}^{OBV}$ are obtained from 48 forward spread indicators MA^{fs} and 15 trading volume-based indicators MA^{OBV} , spanning the period 1964:01 - 2016:12. $PC1 - PC3$, $\hat{f}_{1,t}^{fs} - \hat{f}_{3,t}^{OBV}$, and $PC1 - PC5$ represent the model including the first three PCs of yields, the technical indicator PCs, and the first five PCs of yields, respectively. The top panel reports the average utility gains for the original sample period; the lower panel shows average utility gains for the later sample period.

10 Appendix

Table 12: Statistical inference of technical indicators with binary trading signals

	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>	$\hat{f}_{1,t}^{fs}$	$\hat{f}_{2,t}^{fs}$	$\hat{f}_{3,t}^{fs}$	$\hat{f}_{1,t}^{OBV}$	$\hat{f}_{2,t}^{OBV}$	$\hat{f}_{3,t}^{OBV}$	$\hat{f}_{1,t}^{MOM}$	$\hat{f}_{2,t}^{OBV}$	$\hat{f}_{3,t}^{MOM}$	Wald
<i>Original sample: 1964-2003</i>													
Coefficient	0.125	1.218	-4.348	1.163	-0.941	0.165	0.784	-1.231	-0.835	-0.523	0.600	-0.728	
HAC <i>p</i> -value	0.232	0.050	0.010	0.000	0.018	0.753	0.031	0.119	0.151	0.240	0.315	0.148	0.001
Bootstrap <i>p</i> -value				0.008	0.089	0.800	0.090	0.188	0.210	0.325	0.390	0.221	0.174
IM <i>q</i> = 8	0.030	0.038	0.921	0.847	0.107	0.944	0.248	0.256	0.806	0.129	0.748	0.542	
IM <i>q</i> = 16	0.003	0.145	0.454	0.457	0.395	0.740	0.198	0.182	0.301	0.366	0.209	0.702	
<i>Estimated size of tests</i>													
HAC				0.124	0.158	0.130	0.124	0.098	0.085	0.105	0.086	0.098	0.471
Bootstrap				0.066	0.054	0.056	0.054	0.053	0.054	0.057	0.050	0.051	0.052
IM <i>q</i> = 8				0.074	0.065	0.074	0.069	0.064	0.059	0.062	0.060	0.059	
IM <i>q</i> = 16				0.070	0.063	0.067	0.062	0.055	0.0539	0.063	0.056	0.062	
<i>Estimated power of tests</i>													
HAC				0.802	0.413	0.142	0.676	0.347	0.237	0.243	0.192	0.236	0.962
Bootstrap				0.673	0.221	0.062	0.543	0.244	0.164	0.155	0.129	0.157	0.666
IM <i>q</i> = 8				0.859	0.694	0.643	0.828	0.635	0.593	0.647	0.655	0.735	
IM <i>q</i> = 16				0.847	0.694	0.531	0.810	0.734	0.677	0.656	0.740	0.742	
<i>Later sample: 1985-2016</i>													
Coefficient	0.160	0.234	2.279	1.243	-0.651	0.442	0.022	0.114	-0.055	0.053	0.377	-0.772	
HAC <i>p</i> -value	0.129	0.760	0.311	0.002	0.260	0.375	0.956	0.849	0.881	0.906	0.432	0.268	0.012
Bootstrap <i>p</i> -value				0.022	0.379	0.501	0.967	0.876	0.899	0.927	0.520	0.364	0.430
IM <i>q</i> = 8	0.004	0.011	0.903	0.172	0.233	0.001	0.494	0.177	0.807	0.420	0.313	0.798	
IM <i>q</i> = 16	0.000	0.171	0.602	0.295	0.315	0.323	0.129	0.269	0.144	0.705	0.740	0.927	

Table 12 shows bootstrap results of forecasting yearly excess bond returns (averaged across two through five years) with the level, slope and curvature of the yield curve, the first three technical PC factors of the three technical indicators $\hat{f}_{i,t}^{fs}$, $\hat{f}_{i,t}^{OBV}$ and $\hat{f}_{i,t}^{MOM}$, as well as a binary trading signal. *p*-values of the HAC *t*-test are obtained using Newey-West standard errors with 18 lags. The column "Wald" shows *p*-values for the χ^2 test that all technical PC factors have no predictive power. I also display results for the IM-test, designed by Ibragimov and Müller (2010), for $r = 8$ and 16 subsamples. Estimates of the size and power are shown in panel *Size* and *Power*, respectively. These estimates are obtained from the Monte Carlo simulation described in Bauer and Hamilton (2018). *p*-values below 5% are highlighted in bold.

Table 13: Statistical inference of technical indicators with non-binary trading signals

	$PC1$	$PC2$	$PC3$	$\hat{f}_{1,t}^{fs}$	$\hat{f}_{2,t}^{fs}$	$\hat{f}_{1,t}^{OBV}$	$\hat{f}_{2,t}^{OBV}$	$\hat{f}_{1,t}^{MOM}$	$\hat{f}_{2,t}^{MOM}$	Wald
<i>Original sample: 1964-2003</i>										
Coefficient	0.179	2.170	-5.461	-0.122	-0.879	3.654	-47.243	0.871	-0.423	
HAC p -value	0.125	0.000	0.003	0.522	0.345	0.374	0.000	0.057	0.186	0.000
Bootstrap p -value				0.663	0.539	0.495	0.008	0.173	0.350	0.197
IM $q = 8$	0.000	0.377	0.939	0.659	0.311	0.165	0.827	0.808	0.720	
IM $q = 16$	0.000	0.037	0.170	0.246	0.898	0.062	0.393	0.198	0.883	
<i>Estimated size of tests</i>										
HAC				0.197	0.186	0.158	0.135	0.162	0.168	0.709
Bootstrap				0.050	0.050	0.051	0.048	0.055	0.053	0.044
IM $q = 8$				0.072	0.067	0.075	0.070	0.062	0.063	
IM $q = 16$				0.064	0.061	0.068	0.064	0.055	0.058	
<i>Estimated power of tests</i>										
HAC				0.194	0.240	0.239	0.729	0.282	0.169	0.908
Bootstrap				0.051	0.091	0.099	0.577	0.132	0.058	0.137
IM $q = 8$				0.524	0.682	0.591	0.854	0.621	0.548	
IM $q = 16$				0.589	0.667	0.635	0.892	0.649	0.593	
<i>Later sample: 1985-2016</i>										
Coefficient	0.141	0.505	1.592	191.191	381.096	-0.047	-0.007	-252.203	19.365	
HAC p -value	0.163	0.574	0.395	0.000	0.021	0.944	0.998	0.021	0.840	0.000
Bootstrap p -value				0.005	0.083	0.962	0.998	0.063	0.865	0.148
IM $q = 8$	0.002	0.153	0.242	0.003	0.039	0.029	0.899	0.252	0.058	
IM $q = 16$	0.014	0.175	0.184	0.487	0.024	0.247	0.898	0.127	0.209	

Table 13 shows bootstrap results of forecasting yearly excess bond returns (averaged across two through five years) with the level, slope and curvature of the yield curve, the first three technical PC factors of the three technical indicators $\hat{f}_{i,t}^{fs}$, $\hat{f}_{i,t}^{OBV}$ and $\hat{f}_{i,t}^{MOM}$, and using a non-binary trading signal. p -values of the HAC t -test are obtained using Newey-West standard errors with 18 lags. The column "Wald" shows p -values for the χ^2 test that all technical PC factors have no predictive power. I also display results for the IM-test, designed by Ibragimov and Müller (2010), for $r = 8$ and 16 subsamples. Estimates of the size and power are shown in panel *Size* and *Power*, respectively. These estimates are obtained from the Monte Carlo simulation described in Bauer and Hamilton (2018). p -values below 5% are highlighted in bold.