International Review of Economics and Finance 46 (2016) 10–26

Contents lists available at ScienceDirect

International Review of Economics and Finance

journal homepage: www.elsevier.com/locate/iref

Do economic variables improve bond return volatility forecasts?

Shih-Wei Chao

Department of Money and Banking, National Chengchi University, No.64, Sec.2, Zhinan Road, Taipei 116, Taiwan

ARTICLE INFO

Article history:
Received 23 March 2015
Received in revised form
24 November 2015
Accepted 9 August 2016
Available online 13 August 2016

JEL classification:
C22
G12
G17

Keywords:
Bond return volatility
Predictive ability
Forecast combination
Forecast performance decomposition

ABSTRACT

This paper explores whether various economic variables improve monthly bond return volatility forecasts using the 1963–2012 data. In-sample analysis indicates that stock return or Federal Funds rate difference Granger causes bond volatility of all maturities. The forecasting ability of other variables mainly appears at the short end of the term structure or during the relatively turbulent time. Out-of-sample analysis suggests little evidence of forecast improvement, though forecast combination does improve the performance. Decomposing the out-of-sample forecasts indicates that the poor performance is primarily attributed to overfitting, and variable reduction by principal components does not change the results.

1. Introduction

The average daily trading volume of the US Treasury securities is 545.4 billion dollars in 2013, which is 67% of the volume in all US bond markets and much greater than the 129.68 billion dollars for US stock markets.[1] As one of the largest financial markets in the world, its volatility has rich implications for portfolio evaluation, risk management and policy designation. Empirical analysis for the stock market suggests that return volatility is time-varying and thus more or less predictable. If some predictive relationship also exists in the US Treasury securities, this information would be valuable to bond market watchers. When forecasting bond return volatility, it would be important to investigate the role of macroeconomic and financial indicators for the following reasons. First, the investment decision of US Treasury securities is mostly affected by the macroeconomic condition. Thus the low-frequency variation of bond returns would be particularly important for long-term institutional investors since their decision is usually based on the trade-off between return and risk over a long horizon. Second, asset return volatility is usually high in recessions and periods of high inflation. If bond return volatility also displays a countercyclical pattern, business cycle drivers might contain useful information for predicting bond return volatility. Third, many asset pricing models emphasize the role of fundamentals in the dynamics of returns and risks. The empirical evaluation of the predictive relationship of economic variables and bond return volatility provides stylized facts to

E-mail address: swchao@nccu.edu.tw

[1] http://www.sifma.org/research/statistics.aspx.

http://dx.doi.org/10.1016/j.iref.2016.08.001
1059-0560/

examine the empirical performance of these models. Finally, past studies of volatility comprehensively address the time series properties while left a minor role to the economic environment. As a result, how economic activities and financial indicators signal the dynamics of bond return volatility is still a question of interest.

This paper evaluates whether various economic variables significantly improve the monthly return volatility forecasts for the US Treasuries. Despite the importance of US Treasury markets, the link between low-frequency bond return volatility and the state of the economy receives relatively little attention. As a result, this paper makes at least two contributions to the literature. First, existing studies on bond return volatility prediction focus on a particular maturity or relatively short sample period. This study instead investigates a wide range of maturities, a relatively longer 1963–2012 sample and various specification of forecasting models. The results provide a more comprehensive picture of how the forecasting ability of various economic variables varies with bond maturity and over time. This study also evaluates the out-of-sample performance of several forecast combination schemes. The results would illuminate whether this method significantly reduces variation and bias for bond return volatility forecasts. Second, in-sample and out-of-sample prediction often reach different conclusions in the literature. Although the reliability of two approaches is not the focus of this study, understanding their discrepancy might shed light on how to improve the forecasting model. For example, if undesirable out-of-sample performance is due to changing forecasting ability over time, it would be important to investigate whether this instability is related to structural breaks in parameter estimates or variations of forecasting variables. If the problem is overfitting, excluding some less relevant predictors could lead to better results. This paper provides statistical evidence to identify the primary source of the difference between in-sample and out-of-sample bond return volatility forecasts.

Following Andersen, Bollerslev, Diebold, and Ebens (2001); Andersen, Bollerslev, Diebold, and Labys (2001, 2003), bond return volatility in this paper is measured by realized volatility, which is the sum of squared intra-period returns. The literature has established nice statistical properties for realized volatility under some regular conditions, so its prediction can be proceeded with simple econometric methods. The forecasting model is a linear predictive regression including appropriate lags of bond return volatility and various economic variables as regressors. Although many authors employ different versions of GARCH and successfully describe the salient facts of volatility, they are primarily interested in modeling the clustering, fat-tailed distribution and high-frequency variation of volatility. Because this paper focuses on the role of economic variables in predicting monthly bond return volatility, linear regression is a flexible framework for this purpose. The bond return volatility series are constructed from the CRSP Fixed Term Indices file. Forecasting variables include measures of output growth and its volatility, employment growth, inflation and its volatility, stock return and its volatility, stock market liquidity, default spread, long-short yield spread, and movements in risk-free rate and Federal Funds rate. The predictive regression nests the autoregressive specification as the benchmark when comparing volatility forecasts.

The analysis of the full sample period data indicates that stock return or movements in the Federal Funds rate Granger causes bond return volatility of various maturities. A decline in stock return or a rising Federal Funds rate predicts higher bond return volatility, while the significance decreases in bond maturity. Subsample results suggest that economic variables tend to improve the in-sample fit during relatively turbulent periods, such as sustained high inflation in the 1970s or a series of financial crisis during the late 1990s and early 2000s. The largest increment of R^2 relative to the benchmark is below 3%, which lacks impressive economic significance. On the other hand, the out-of-sample evaluation displays weaker evidence of forecasting ability. Most augmented models with economic variables do not beat the benchmark. In fact, some of these models produce very poor forecasts, particularly when the model contains a large number of predictors. Following Rapach, Strauss, and Zhou (2010), this paper also evaluates the out-of-sample performance of several forecast combination schemes. The results indicate that ensemble bond return volatility forecasts appear to outperform the benchmark in some circumstances. The null of equal predictive ability can be rejected in favor of smaller forecast errors for some combined forecasts for 1- and 10-year Treasuries. The combined forecasts also tend to perform better in the relatively unsettled periods, but the improvement of out-of-sample R^2 relative to the benchmark is below 2% .

To explain the difference between in-sample and out-of-sample forecasting performance, this study follows Rossi and Sekhposyan (2011) to decompose the out-of-sample mean forecast error into three asymptotically uncorrelated components, namely forecast instability, predictive content and overfitting. For those significantly worse volatility forecasts, the tests often rejects the null of no overfitting but fails to reject the null of no forecast instability or no lack of predictive content. To explore whether variable reduction is a remedy to this overfitting problem, the predictors are replaced by several principal components of these economic variables. Unfortunately, using these principal components rarely improves the out-of-sample performance relative to the benchmark. In sum, these economic variables provide little useful information in addition to the benchmark autoregression model when making out-of-sample bond return volatility forecasts, particularly for intermediate and long maturities.

This paper relates to the literature of the link between low-frequency bond return volatility and macroeconomic variables.[2] Christiansen, Schmeling, and Schrimpf (2012) explore the economic determinants of return volatility for a variety class of assets, including the 10-year Treasury Notes futures. The authors use Bayesian model averaging to determine the specification of forecasting models and rank the predictive ability of each model accordingly. Their results based on the

[2] A strand of literature provides evidence on the effects of macroeconomic news announcements on bond market volatility, such as Bollerslev, Caic, and Song (2000); Brenner, Pasquariello, and Subrahmanyam (2009); Jones, Lamont, and Lumsdaine (1998), among others. These studies mostly focus on analyzing high-frequency (intra-day or daily) variation of bond returns.

1983–2010 sample suggest that stock turnover, default spread and term spread are useful predictors for bond return volatility. Viceira (2012) studies how time-varying bond risk and bond return volatility are connected with the term structure of interest rates using the data of 5-year US Treasuries. His analysis indicates that movements in the short-term interest rate and the spread between long and short yields tend to be positively related with future bond return volatility, but the latter is usually not statistically significant. Since the above studies focus on a certain term to maturity, a relatively short sample period or a specific set of forecasting variables, this paper provides more extensive evidence on bond return volatility of different maturities and historical periods. This paper also relates to studies of stock return volatility prediction using linear models. For example, Paye (2012) finds that measures of macroeconomic uncertainty, expected stock returns and credit conditions Granger cause stock return volatility, particularly around the early stage of recessions. Exploiting information of macroeconomic variables is not conducive to out-of-sample prediction, while forecast combination improves the performance. Despite similar methodologies, the current paper further employs a statistical diagnosis to identify the primary source of poor out-of-sample forecasts.

2. Data and empirical methodology

2.1. Bond return volatility

The empirical measure of monthly realized bond return volatility is

$$RV_t = \sqrt{\sum_{i=1}^{N_t} \left(R_{i,t}^{n}\right)^2}, \tag{1}$$

where N_t denotes the number of trading days in month t and $R_{i,t}^{n}$ is the daily bond return of maturity n. This measure of volatility simply involves the sum of squared intra-period returns, so the computation does not require any specific model or distributional assumption. Unlike the measures produced by models of conditional or stochastic volatility, realized volatility is an observable measure rather than a latent variable to be estimated. As discussed in Andersen et al. (2003), the theory of quadratic variation implies that various realized volatilities converge to the unobserved integrated volatility under some regular conditions. These advantages advocate the use of this measure in the literature of various return volatilities. On the other hand, the distribution of realized volatility is right-skewed and leptokurtic. This feature could deteriorate the validity of empirical analysis based on predictive regressions. Fortunately, the logarithmic realized volatility behaves more like a Gaussian process. As a result, the subsequent analysis proceeds with the logarithm of realized volatility $\ln RV_t$.

The data of daily bond returns of maturity 1, 5, 10, and 30 years are provided by the CRSP Fixed Term Indices file from January 2, 1963 to December 31, 2012.[3] Table 1 reports the mean, standard deviation and some autocorrelation coefficients of RV_t. It is clear that long-term bond returns are more volatile since the mean increase from 0.29% for the 1-year Treasuries to 2.81% for the 30-year Treasuries. Bond return volatilities are also persistent. The first-order autocorrelation coefficients range from 0.75 for the 1-year Treasuries to 0.85 for the 30-year Treasuries, and the 12th-order autocorrelations are still around 0.5 to 0.6. Thus the persistence unambiguously increases in maturity and decays slowly. Because bond return volatilities are highly correlated over time, the autoregressive specification serves a benchmark in the subsequent prediction evaluations. Additional economic variables help to predict bond return volatility only when they significantly improve the forecasting performance relative to the benchmark.

The volatilities of US Treasuries also exhibit substantial fluctuations over time. Fig. 1 illustrates the standardized logarithmic realized volatilities of maturity 1, 5, 10, and 30 years. Bond return volatilities appeared to be somewhat turbulent before the mid-1980s and then became relatively stable until the early 2000s. This pattern is roughly consistent with the transition of the US economy from Great Inflation to Great Moderation. The series also have spikes around the deep recession in the early 1980s, the stock market crash in October 1987 and the financial crisis in 2008. In addition, Table 2 provides a comparison of bond return volatility between the recession and non-recession periods. The countercyclical feature of asset return volatility also appears in bond markets: volatilities are clearly higher during recessions. The median of standardized bond return volatility can be as high as 0.82 during recessions, while the values are not far from zero during normal times. Although bond returns tend to be more volatile during recessions and periods of turbulent financial markets, this contrast is more evident on the short end of the term structure while less remarkable on the long end. This countercyclical pattern of bond return volatilities motivates the use of various economic and financial variables in their prediction, but the predictability might vary across maturities.

[3] The CRSP Fixed Term Indices file contains various statistics for constant maturity coupon bonds with maturities of 1, 2, 5, 7, 10, 20 and 30 years starting from June 14, 1961. This paper focuses on maturities of 1, 5, 10 and 30 years, which represent the shortest available, the typical intermediate, the indicative long-term and the longest available maturity, respectively. The sample period starts from January 2, 1963 to match the available data of all forecasting variables. Results of other maturities are available upon request.

Table 1
Descriptive statistics for realized bond return volatilities.

	1-Year	5-Year	10-Year	30-Year
Mean	0.2879	1.1777	1.7991	2.8097
Std. Dev.	0.2124	0.7818	1.0531	1.6807
AC(1)	0.7531	0.7623	0.8081	0.8459
AC(2)	0.7107	0.7186	0.7742	0.8081
AC(6)	0.6270	0.5852	0.6716	0.7004
AC(12)	0.4966	0.4795	0.5991	0.6109
ADF	0.0000	0.0000	0.0000	0.0000
PP	0.0000	0.0000	0.0000	0.0000

Note: The table reports summary statistics of realized bond volatilities. AC(n) means the n-th order autocorrelation coefficient. ADF and PP report the MacKinnon approximate p-value for augmented Dickey-Fuller test and Phillips-Perron test for unit root, respectively.

2.2. Forecasting variables

As discussed below, the forecasting variables are suggested by various empirical evidence and theoretical results in the literature.

- Industrial production growth: This variable is measured by the growth rate of Industrial Production Index.
- ISM index: It is measured by the first-order difference of the Institute for Supply Management (ISM) Manufacturing New Orders Index. This variable might contain information for future industrial production.
- Employment condition: This variable is measured by the growth rate of Total Nonfarm Payroll (TNP), where TNP is the number of workers in the US economy.
- Inflation: The conventional measure is the growth rate of Consumer Price Index (CPI). I also consider the growth rate of Personal Consumption Expenditure (PCE) as a measure of inflation because the PCE index covers a wide range of household spending and closely monitored by the Fed.
- Stock market return: Stock return is measured by the return on S&P 500 index from CRSP.
- Stock return volatility: Stock volatility is measured by the realized variance of S&P 500 index, which is the same as Goyal and Welch (2008).

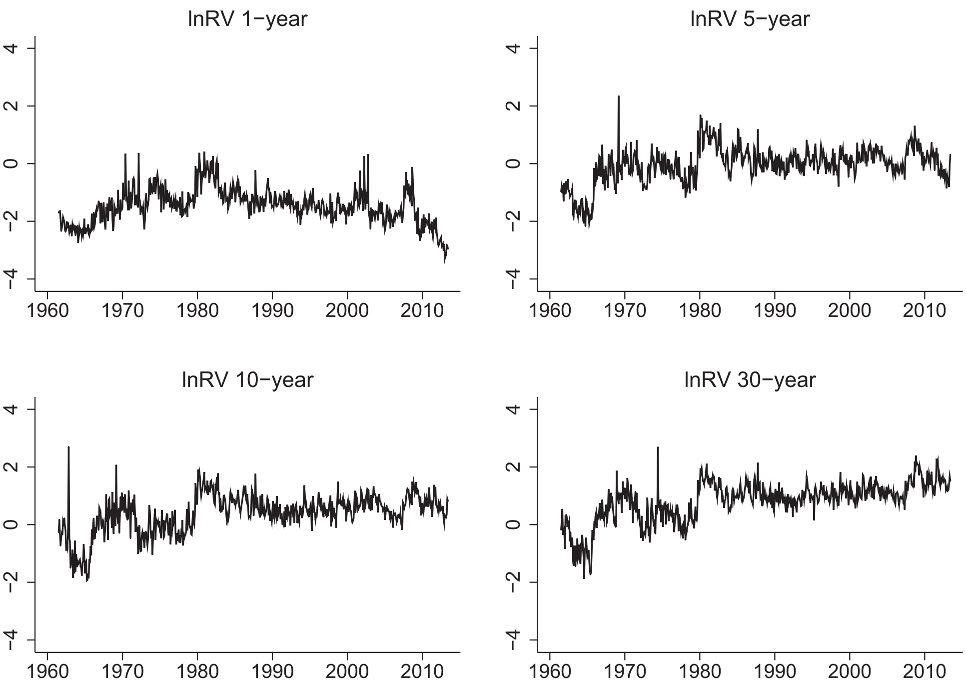

Fig. 1. Monthly log bond return volatility. *Note*: The figure displays the standardized time series of monthly log bond return volatility for the 1963–2012 period.

Table 2
Bond return volatilities in recession and non-recession periods.

	Recession			
	1-Year	5-Year	10-Year	30-Year
Mean	0.9266	0.7257	0.5673	0.5877
Median	0.8248	0.6978	0.7280	0.6443
Std. Dev.	0.9830	0.8303	0.9239	0.8551
	No recession			
	1-Year	5-Year	10-Year	30-Year
Mean	−0.1488	−0.1165	−0.0911	−0.0943
Median	−0.1428	0.0213	0.0868	0.1142
Std. Dev.	0.9203	0.9762	0.9824	0.9900

Note: The table reports summary statistics of standardized realized bond volatilities for US recession periods (83 months in our sample) and non-recession periods, respectively.

- Stock market liquidity: This is measured by the Pástor and Stambaugh (2003) factor, which is available from Stambaugh's website.
- Default spread: This is measured by the difference of yield on Baa-rated and Aaa-rated corporate bonds.
- Term spread: This is the difference between the yield on long-term government bonds and the Treasury bill rate.
- Risk-free rate difference: This is measured by the first-order difference of the Treasury bill rate.
- Federal Funds rate difference: The first-order difference of the Federal Funds rate is a proxy for the change of monetary policy.
- Volatilities of output growth and inflation: The estimation of the standard deviation of industrial production growth, CPI and PCE inflation follows Schwert (1989).

Industrial production growth, ISM index, employment growth and measures of inflation are considered as forecasting variables primarily due to the potential link between bond return volatilities and business cycles. Exploiting information in these macroeconomic indicators could improve the prediction of countercyclical volatilities. The predictability of bond risk premium also motivates the use of these macroeconomic variables. Numerous asset pricing studies suggest that asset return fluctuations over time is closely related to time-varying aggregate price of risk in the risk premium. Ludvigson and Ng (2009) study a comprehensive list of macroeconomic variables and find that factors relating to real economic activities and inflation have important forecasting power for US bond risk premium. As a result, measures of productivity and inflation could also forecast bond return volatility. On the other hand, Bansal and Shaliastovich (2013) find that excess bond returns rise with uncertainty about expected inflation and fall with uncertainty about real consumption growth. This result implies that volatilities of these macroeconomic variables could also forecast bond return volatilities.

The cyclical behavior of bond return volatilities also motivates the use of term structure variables since the yield curve contains information for future economic activities. A long strand of literature suggests that term spread forecasts GDP growth, while Ang, Piazzesi, and Wei (2006) argue that the short-term interest rate has more significant forecasting power. If these term structure factors forecast business cycles, they could also forecast bond return volatilities. Thus term spread and risk-free rate difference are included as forecasting variables. In addition, yield curve is sensitive to changes in monetary policy. Thus Federal Funds rate difference is also a candidate since it represents the movements in the policy instrument monitored by the Fed. The tight link between monetary policy and inflation also motivates the use of this variable in bond return volatility prediction.

The remaining forecasting variables concern the link between bond and other financial markets. Empirical analysis indicates the modest positive correlation of stock and bond returns, though this correlation displays considerable fluctuations over time. Fleming, Kirby, and Ostdiek (1998) show that portfolio hedging across markets establishes a tight link between stock and bond volatilities. Goyenko and Ukhov (2009) provide the evidence on the liquidity link between stock and bond markets. These findings suggest that aggregate stock return, volatility and liquidity could have nontrivial effects on bond return volatilities. On the other hand, lessons from various financial crisis highlights the importance of flight to quality. Funds often flow from risky assets to securities with little default risks during episodes of financial turmoils. Brunnermeier and Pedersen (2009) show that tightening funding liquidity during a crisis could force investor to rapidly sell their positions of risky assets. This behavior quickly depletes market liquidity and affects volatilities of both risky and safe assets. Fontaine and Garcia (2012) find that the decrease in funding liquidity lowers the risk premia on US Treasury bonds but raises the risk premia on other debt instruments. As a result, default spread is a potential predictor.

All of the forecasting variables are sampled at monthly frequency. Table 3 presents their summary statistics and the *p*-values of unit root tests. Although some inflation measures, default spread and term spread are fairly persistent, the hypothesis of unit root is rejected at the usual significance level for all variables. As a result, it is safe to conclude that all these forecasting variables are stationary time series.

Table 3
Descriptive statistics for forecasting variables.

	Mean	Std. Dev.	AC(1)	AC(6)	ADF	PP
IP growth	0.0023	0.0075	0.3294	0.1197	0.0000	0.0000
ISM index	−0.0167	3.7887	−0.0443	−0.0155	0.0000	0.0000
Employment growth	0.0015	0.0026	0.6064	0.2210	0.0000	0.0000
CPI inflation	0.0034	0.0035	0.6112	0.2659	0.0000	0.0000
PCE inflation	0.0030	0.0025	0.7131	0.5491	0.0000	0.0000
IP growth volatility	0.0048	0.0015	0.1872	−0.1118	0.0000	0.0000
CPI inflation volatility	0.0018	0.0006	0.5890	0.5497	0.0000	0.0000
PCE inflation volatility	0.0012	0.0004	0.5852	0.3971	0.0000	0.0000
Stock return	0.0088	0.0433	0.0402	−0.0598	0.0000	0.0000
Stock volatility	0.0403	0.0243	0.6682	0.3492	0.0000	0.0000
Stock liquidity	−0.0306	0.0631	0.1326	0.1164	0.0000	0.0000
Default spread	0.0104	0.0047	0.9668	0.7267	0.0203	0.0028
Term spread	0.0152	0.0129	0.9576	0.7034	0.0057	0.0013
RF rate difference	−0.0004	0.0004	0.3347	−0.1719	0.0000	0.0000
FF rate difference	−0.0046	0.0055	0.4030	−0.0597	0.0000	0.0000

Note: The table reports summary statistics of various forecasting variables. AC(n) means the n-th order autocorrelation coefficient. ADF and PP report the MacKinnon approximate p-value for augmented Dickey-Fuller test and Phillips-Perron test for unit root, respectively.

2.3. Forecasting model

The bond return volatility forecasts are generated from the predictive regression

$$\ln RV_t = \alpha + \sum_{j=1}^{p} \rho_j \ln RV_{t-j} + \boldsymbol{\beta}' \mathbf{X}_{t-1} + u_t, \tag{2}$$

where p is the appropriate lags for the logarithmic bond return volatility, $\mathbf{X}_{t-1}$ is a set of forecasting variables and $\boldsymbol{\beta}$ is the corresponding vector of estimated coefficients. The choice of p is based on the Schwarz Bayesian information criterion (SBIC). All variables are standardized prior to estimation so that the economic significance of different forecasts can be easily compared. As a result, any estimated coefficient is interpreted as the change in standard deviation of log bond return volatility in association of the unit change in standard deviation of the forecasting variable. To explore the forecasting ability of different variables, the vector of predictors $\mathbf{X}_{t-1}$ may be univariate or a multivariate specification.[4] As reported in Tables 1 and 3, the hypothesis of unit root can be safely rejected at the 5% significance level for all variables. Thus the usual statistical inferences are appropriate for these predictive regressions. The standard errors for inference are Newey-West heteroskedasticity and autocorrelation consistent estimates.

3. In-sample analysis

The full-sample estimation results are presented in Table 4. First, the predictability of most economic variables displays considerable variation across maturities. Two measures of inflation, volatility of industrial production growth, stock return, default spread, term spread, risk-free rate difference or Federal Funds rate difference Granger causes 1-year bond return volatility. However, many of them fail to forecast volatility of longer maturities. In contrast, variables associated with output and productivity sometimes Granger cause bond return volatility of intermediate and long maturities but not at the short end of the term structure. Industrial production growth forecasts volatility of 5- and 10-year Treasuries. The ISM index also shows forecasting ability when maturity extends to 10 years and above. Second, stock return and movements in the Federal Funds rate are the most prominent forecasting variables since they are statistically significant across the term structure. Holding other things constant, bond returns are predicted to be less volatile when stock return is high or Federal Funds rate tends to decline. Finally, a large set of economic variables shows joint forecast ability for volatility of maturity 10 years and below. This specification, however, simply improves the model fit by a few percent and the increase in R^2 unanimously declines in bond maturity. As a result, increasing the number of economic variables in the forecasting model is not particularly useful for bond return volatility prediction.

Although stock return appears to Granger cause bond return volatility across maturities, some studies indicate that excess stock return could be driven by bond volatility. For example, Mueller, Vedolin, and Yen (2011) show that three principal components of the bond variance risk premium explain about 9% of the variation of future stock excess returns. As

[4] The multivariate specification includes all forecasting variables except employment growth, risk-free rate difference, PCE inflation and its volatility due to the concern of multicollinearity.

Table 4
Full-sample predictive regressions.

	1-Year	5-Year	10-Year	30-Year
IP growth	−0.0043	−0.0498**	−0.0473*	−0.0289
ISM index	−0.0252	−0.0128	−0.0428*	−0.0328*
Employment growth	0.0139	−0.0323	−0.0288	−0.0258
CPI inflation	0.0815***	0.0329	0 .0212	0.0040
PCE inflation	0.0997***	0.0441	0 .0215	0.0054
IP growth volatility	−0.0506*	−0.0233	−0.0347	−0.0297
CPI inflation volatility	−0.0406	0.0019	−0.0013	0 .0019
PCE inflation volatility	−0.0083	0.0127	0.0110	−0.0006
Stock return	−0.0881***	−0.0698***	−0.0446**	−0.0416*
Stock volatility	−0.0302	−0.0018	0.0134	0.0259
Stock liquidity	0.0058	0.0004	−0.0008	−0.0166
Default spread	−0.0648*	0.0133	0.0154	−0.0004
Term spread	−0.0711***	−0.0251	−0.0244	−0.0298
RF rate difference	0.1061***	0.0766**	0.0387	0.0213
FF rate difference	0.1035***	0.0855***	0.0540**	0.0430**
Benchmark R^2	0.6484	0.6424	0.7082	0.7554
Large				
ΔR^2	0.0257***	0.0178***	0.0109*	0.0068

Note: The table reports the estimated coefficients of various bond return volatility predictive regressions for the full sample period 1963M1–2012M12. The superscripts ***, ** and * indicate the statistical significance at 1%, 5% and 10%, respectively. "Large" means all forecasting variables are included in the predictive regression except employment growth, risk-free rate difference, PCE inflation and its volatility since they are highly correlated with some other forecasting variables. Benchmark R^2 reports the R^2 for the benchmark autoregression model. ΔR^2 is the increase in R^2 by the large model relative to the benchmark, and the superscripts ***, ** and * indicate that the null of no joint significance can be rejected at 1%, 5% and 10%.

a result, it would be important to investigate whether the significance of stock return is simply a consequence of simultaneous causality. The investigation includes regressing stock return on the residual from the bond return volatility predictive regression, regressing stock return on its past values and lags of bond return volatility, and estimating a VAR system including return, volatility and liquidity of stock market and bond return volatility to control the effects of other stock market variables. As shown in Table 5, the residual from the predictive regression does not forecast stock return. Either linear regression or VAR indicates that bond return volatility does not Granger cause stock return. This conclusion is robust across the entire term structure and not sensitive to the order of variables in the VAR.[5] As a result, stock return is unlikely an endogenous regressor in the forecasting model, and the forecasting ability of stock market performance on bond return volatility is not subject to the simultaneous causality problem.

The significance of Federal Funds rate difference appears to reflect the role of monetary policy in determining bond return volatility since this is the policy instrument closely monitored by the Fed. However, movements in the Federal Funds rate include the movements of expected Federal Funds rate (perhaps based on a monetary policy rule) as well as unexpected monetary policy shocks. If the forecasting power is actually driven by the former, the predictive ability of movements in the Federal Funds rate difference could be a consequence of endogeneity problem. To understand the source of its forecasting ability, difference in the Federal Funds rate is decomposed into expected movements and unexpected policy shocks, where the latter is estimated by the least square residuals from regressing Federal Funds rate difference on inflation and industrial production growth. Each component then serves as a forecasting variable in Eq. (2) and the associated estimates are summarized in Table 6. It is clear that the coefficient in the unexpected policy shocks model is comparable to that in the movements in the Federal Funds rate model, and this component forecasts bond return volatility across all maturities. In contrast, the expected movements in the Federal Funds rate is never a significant predictor in any case. As a result, the significance of movements in the Federal Funds rate is likely driven by unexpected monetary policy shocks.

Despite the in-sample significance of some variables, their forecasting ability mostly concentrates on short-term Treasuries. Even if stock return or movements in the Federal Funds rate forecasts bond return volatility across the entire term structure, their statistical significance also declines in bond maturity. Because the persistence of bond return volatility increases in term to maturity, the dynamics of long-term bond return volatility tend to be more dependent on its past history. As a result, the predictability enhanced by additional explanatory variables would be limited if the forecasting model is controlled for appropriate lags of bond return volatility. On the other hand, the economic significance of these economic variables is a question. Note that the largest estimated coefficient is simply around 0.1 in Table 4. Since all variables are standardized prior to estimation, one standard deviation change in a forecasting variable at best moves future bond return volatility by one-tenth of the standard deviation. Although it is not uncommon to have small estimates in predictive regressions, these coefficients still reflect limited enhancement of the predictive ability by economic variables. In

[5] The same VAR estimation also confirms the Granger causality from stock return to bond volatility of short and intermediate maturities.

Table 5
Tests of causality from bond return volatility to stock return.

	Residual	Reverse	VAR
1-Year	0.592	0.939	0.370
5-Year	0.673	0.653	0.712
10-Year	0.995	0.664	0.477
30-Year	0.333	0.809	0.889

Note: The table reports the *p*-values of testing whether bond return volatility causes stock return. "Residual" means the regression of stock return on the residual from the forecasting model. "Reverse" means the regression of stock return on lagged bond return volatilities. "VAR" means the vector autoregression containing bond return volatility, stock return, stock volatility and stock liquidity. The null hypothesis of the test is that bond return volatility does not Granger cause stock return.

Table 6
Predictability of components in the Federal Funds rate difference.

	Unexpected policy shock		Expected FF rate difference	
	Coefficient	*p*-Value	Coefficient	*p*-Value
1-Year	0.1052	0.000	0.0722	0.451
5-Year	0.1034	0.000	−0.1172	0.198
10-Year	0.0704	0.028	−0.1261	0.154
30-Year	0.0549	0.023	−0.0882	0.269

Note: The table reports the estimated bond return volatility predictive regressions using unexpected monetary policy shock or expected movements in the Federal Funds rate as the predictor, respectively.

addition, Table 4 also shows that a large set of forecasting variables simply increases the R^2 by less than 3%. As a result, these economic variables appear to provide limited forecasting information in addition to the benchmark model.

Because the significance of forecasting variables might change over time, the predictive regressions are also estimated for three shorter horizons. The first one covers 1963M1–1983M12, which includes several years of stagflation and productivity slowdown. The subsequent 1984M1–1997M6 and 1997M7–2012M12 periods primarily characterize the Great Moderation, but the former is possibly a more tranquil regime since the latter includes several events negatively affecting the world economy.[6] As displayed in Table 7, the statistical significance of each forecasting variables has some variations over different historical periods. Two measures of inflation, stock return, risk-free rate difference or movements in the Federal Funds rate forecasts bond return volatility across the entire term structure during the 1963M1–1983M12 horizon. Term spread is also a useful predictor in this period except for the volatility of 5-year Treasuries. In contrast, the evidence of predictability is weaker in the relatively stable 1984M1–1997M6 horizon. This can be detected from fewer significant estimates and generally minor economic significance in terms of smaller estimated coefficients. As a result, these variables tend to have stronger forecasting ability in periods of relatively turbulent economic environment.

4. Out-of-sample analysis

If the data support a predictive relationship among bond return volatility and several economic variables in a certain historical period, economists and practitioners might wonder whether this relationship still holds in the future. More explicitly, the question of interest would be whether the volatility forecasts from Eq. (2) with $\beta \neq 0$ outperform the benchmark beyond the current sample. However, evaluating the performance of these forecasts in the future is not feasible in real time since the realizations are not available at the time of prediction. As a result, the common practice is to conduct a "pseudo" out-of-sample experiment. The data are split into an estimation sample and a holdout sample. The historical relationship is estimated using the observations in the estimation sample. The forecasts for the holdout sample are constructed from the estimated relationship as if the data in the holdout sample were unavailable at the time of doing predictions. The out-of-sample performance can be examined by analyzing the difference between forecasts and data in the holdout sample. Following this procedure, the out-of-sample analysis starts from estimating various specifications of Eq. (2) using a fixed window of 240 months or a recursive scheme with the initial window of same length. The out-of-sample forecasts are then

[6] These events include but not limit to the Asian financial crisis starting in July 1997, the Russian crisis in 1998, the burst of dot com bubble in 2000 and various economic crises in several emerging markets during 1998–2003. In addition, prices of many raw materials steadily rose since 2002 and reached historical highs before the 2008 crisis. Hence this period is in contrast with the regime of low and stable inflation in the 1990s.

Table 7
Subsample results.

	1963M1–1983M12				1984M1–1997M6				1997M7–2012M12			
	1-Year	5-Year	10-Year	30-Year	1-Year	5-Year	10-Year	30-Year	1-Year	5-Year	10-Year	30-Year
IP growth	−0.01	−0.03	−0.04	−0.01	−0.12***	−0.10*	−0.06	−0.06	0.02	−0.07*	−0.06**	−0.05**
ISM index	−0.03	0.01	−0.04	−0.02	−0.01	−0.01	−0.01	0.00	−0.06	−0.06*	−0.08***	−0.08***
Employment growth	−0.03	−0.04	−0.03	−0.02	0.01	0.06	0.08	0.04	0.08	−0.05	−0.08**	−0.06*
CPI inflation	0.17***	0.11***	0.10**	0.09**	0.09	0.05	0.04	0.02	0.00	−0.02	−0.03	−0.05**
PCE inflation	0.18***	0.13***	0.11**	0.11***	0.05	0.01	−0.02	−0.02	0.03	−0.00	−0.04	−0.08***
IP growth volatility	−0.08	−0.01	−0.05	−0.05*	−0.09**	−0.10**	−0.07**	−0.07***	−0.10**	−0.01	0.02	0.05
CPI inflation volatility	−0.01	0.05	0.03	0.03	−0.04	−0.10**	−0.05	−0.04	−0.07*	−0.02	−0.01	0.01
PCE inflation volatility	0.03	0.04	0.02	0.00	0 .05	−0.02	0.02	0.04	−0.04	−0.01	−0.00	−0.01
Stock return	−0.16***	−0.13***	−0.07**	−0.07*	−0.03	0.04	0.03	0.02	−0.08	−0.08***	−0.07***	−0.06**
Stock volatility	0.19	−0.00	−0.00	0.00	−0.04	−0.07***	−0.06**	−0.05***	−0.05	0.02	0.08***	0.10***
Stock liquidity	−0.10**	−0.03	−0.01	−0.05	0.09**	0.04	0.03	0.02	0.05	0.01	−0.02	−0.02
Default spread	−0.02	0.03	0.01	−0.01	0.14***	0.02	0.11**	0.03	−0.15***	−0.01	0.07***	0.12***
Term spread	−0.10**	−0.08	−0.08*	−0.09**	0.01	0.05	0.01	−0.02	−0.10**	−0.01	0.05	0.04
RF rate difference	0.10***	0.09**	0.05*	0.04*	0.13**	0.03	0.03	0.02	0.07	−0.10	−0.16***	−0.17***
FF rate difference	0.10***	0.10***	0.07**	0.06**	0.11	0.03	−0.00	0.01	0.13	−0.16*	−0.19**	−0.18**
Benchmark R^2	0.69	0.70	0.74	0.73	0.34	0 .23	0.24	0.23	0.58	0.47	0.41	0.50

Note: The table reports the estimated coefficients of bond return volatility predictive regressions for three subsample periods. Benchmark R^2 reports the R^2 for the benchmark autoregression model. The superscripts ***, ** and * indicate the statistical significance at 1%, 5% and 10%, respectively.

constructed based on these estimates. Their performance is evaluated for the subsequent 1983–2012 period and three shorter horizons, namely 1983–1992, 1993–2002 and 2003–2012.

Forecast accuracy is measured by the sample mean squared forecast error (MSFE). Smaller MSFE for a forecasting model relative to the benchmark might be a signal of its superior predictive ability, and its statistical significance can be examined by the Giacomini and White (2006) test.[7] Let $MSFE_B$ be the sample MSFE for the benchmark model and $MSFE_i$ be the sample MSFE for a forecasting model with economic variables as additional regressors. The associated test statistic for unconditional predictive ability can be written as

$$GW = \frac{MSFE_B - MSFE_i}{\hat{\sigma}_P/\sqrt{P}}, \tag{3}$$

where $\hat{\sigma}_P/\sqrt{P}$ denotes the associated heteroskedastic and autocorrelation robust standard error. The asymptotic distribution of this statistic is $N(0,1)$ under the null of equal predictive ability. On the other hand, the out-of-sample performance of a prediction model can also be gauged by the out-of-sample R^2 statistic proposed by Campbell and Thompson (2008):

$$R^2_{OS} = 1 - \frac{MSFE_j}{MSFE_0}, \tag{4}$$

where $MSFE_j$ denotes the sample MSFE for the model of interest and $MSFE_0$ is the MSFE for the historical bond return volatility. As a result, the improvement of out-of-sample R^2 provided by an augmented model relative to the benchmark can be written as

$$\Delta R^2_{OS} = \left(1 - \frac{MSFE_i}{MSFE_0}\right) - \left(1 - \frac{MSFE_B}{MSFE_0}\right) = \frac{MSFE_B - MSFE_i}{MSFE_0}. \tag{5}$$

The size of the ΔR^2_{OS} statistic delivers the economic significance of forecasting improvement by the model of interest relative to the benchmark.

Tables 8, 9, 10 and 11 display the ΔR^2_{OS} statistics and the results of Giacomini-White test for various volatility forecasts. In contrast with the in-sample results, models with additional economic variables rarely generate superior forecasts for bond return volatility. It is difficult to reject the null of equal predictive ability in most cases, and this conclusion is robust to different bond maturities, estimation schemes or forecasting periods. Even if models with stock return or movements in the Federal Funds rate have prominent in-sample fit for all bond maturities, most of their out-of-sample forecasts do not display superior predictive ability. The poor performance of out-of-sample volatility forecasts is also evident in the size of the ΔR^2_{OS} statistics. Even the largest ΔR^2_{OS} among all the statistically superior forecasts is less than 3%, implying limited forecasting improvement relative to the benchmark.[8] The analysis also examines the out-of-sample performance of the model with many forecasting variables. Unfortunately, most of these forecasts are statistically inferior to the benchmark and the associated ΔR^2_{OS} can be as low as -15%. Thus increasing the number of forecasting variables in predictive regressions appears to amplify the out-of-sample forecasting error.

Although many forecasting models do not provide superior out-of-sample volatility forecasts, the details appear to vary across the term structure. Including some term structure variables, measures of interest rates or the volatility of industrial production growth occasionally produce more accurate volatility forecasts for 1-year Treasuries. For Treasuries of longer maturities, models with additional economic variables hardly provide superior volatility forecasts. Even worse, the hypothesis of equal predictive ability is numerously rejected in favor of the benchmark. This pattern somewhat echoes the in-sample result that the predictive ability is lower for long-term bonds. On the other hand, the out-of-sample performance also fluctuates across different forecasting periods. Most statistically superior volatility forecasts for 1-year Treasuries are found in the 2003–2012 horizon, which features a relatively unsettled environment due to the US subprime mortgage and European sovereign debt crises. Thus the out-of-sample predictive ability is slightly stronger in the relatively turbulent periods for volatility of the short-term Treasuries. Note that this improvement is not very prominent in terms of economic significance since the positive ΔR^2_{OS} statistics are all less than 2%.

Because forecasts from a certain predictive regression might be vulnerable to estimation errors and model misspecifications, combining individual forecasts might reduce forecasting variations and yield more reliable out-of-sample performance. The combined forecasts for volatility $\ln\hat{RV}_{t+1}$ made at time t is defined as

$$\ln\hat{RV}_{t+1} = \sum_{k=1}^{n} w_{k,t}\ln\hat{RV}_{k,t+1}, \tag{6}$$

[7] The question of interest here is the finite sample predictive ability, i.e., the accuracy of the forecasts based on estimated model parameters rather than unknown population parameters. Since the comparison involves nested models, the method proposed by Giacomini and White is powerful and easy to implement. One concern of this test is that it focuses on limited memory forecasting methods. When forecasts are from recursively estimated parameters, the distribution of the test statistic might be nonstandard and the critical values might rely on bootstrap procedures.

[8] The out-of-sample forecasting performance is also evaluated by a 5% VaR back test. Since the probabilities that the VaR fails to cover the actual loss are mostly equal for all models of the same maturity, models with economic variables do not display different forecasting ability relative to the benchmark. I thank the suggestion from an anonymous referee for this exercise and the relevant result is available upon request.

Table 8
Out-of-sample results: one-year bond return volatility.

	1983–2012		1983–1992		1993–2002		2003–2012	
	Roll	Recur	Roll	Recur	Roll	Recur	Roll	Recur
IP growth	−0.43	−0.15	0.86	0.96*	0.22	−0.06	−1.40	−0.61
ISM index	−0.19	−0.00	−0.10	−0.23	−1.00	−0.12	0.03	0.12
Employment growth	−0.44	−0.18	−0.35	0.34	0.07	0.08	−1.02	−0.59
CPI inflation	−0.19	−0.91	1.14	0.37	−0.65	−0.31	−0.49	−2.22
PCE inflation	0.10	−0.31	−1.34	−2.66	1.98	1.91	−0.31	−0.92
IP growth volatility	0.55	0.52*	0.68	1 .12*	−1.49	−0.25	1.88**	1.09*
CPI inflation volatility	0.24	0.23	−0.26	−0.21	−0.03	−0.11	0.69	0.68**
PCE inflation volatility	−0.30	−0.25	0.42	0.63	0.04	0.37	−0.88	−0.96
Stock return	−0.88	−0.60	−7.29	−8.25	2.10	3.98	−1.12	−1.04
Stock volatility	−1.86	−2.07	−14.9	−17.0	−1.35	−0.65	0.26	−0.01
Stock liquidity	−0.32	−0.91*	−4.17	−4.38	−0.85	−1.51	0.79	−0.30*
Default spread	1.00*	0.41	−0.36	0.01	1.53	0.74	1.84	0.67
Term spread	−0.38	0.27	−1.75	−2.44	−1.80	−0.06	0.41	1.38**
RF rate difference	0.36	0.62*	2.33	2.25	−0.25	−0.71	0.36**	1.24***
FF rate difference	0.19	0.43	2.09	1.99	−0.34	0.12	0.03	0.46
Large	−2.24	−3.38	−18.5	−19.7	−1.80	1.73	0.55	−3.59*
Mean	0.31	0.31**	0.24	0.06	0.11	0.72	0.65*	0.41*
Median	0.16	0.20**	0.46	0.31	0.16	0.09	0.20	0.36**
Trimmed mean	0.35**	0.32***	0.86	0.73	0.15	0.36	0.54*	0.42**
MSFE	0.37*	0.37*	0.73	0.58	0.23	0.79*	0.71**	0.48*

Note: The table reports the increase in out-of-sample R^2 relative to the benchmark when the forecasting model includes economic variables. The numbers in the table are their original values multiplied by 100. The superscripts ***, ** and * indicate that the null of equal predictive ability can be rejected at 1%, 5% and 10%, respectively. "Large" means the model with a large set of forecasting variables. "Roll" and "Recur" represent the rolling and recursive estimation scheme, respectively.

Table 9
Out-of-sample results: five-year bond return volatility.

	1983–2012		1983–1992		1993–2002		2003–2012	
	Roll	Recur	Roll	Recur	Roll	Recur	Roll	Recur
IP growth	1.11	0.74	1.53**	1.44**	0.25	−0.41	1.30	0.88
ISM index	−0.09	−0.18	−0.82	−0.53	0.17	−0.12	0.34	0.06
Employment growth	0.29	0.11	0.16	0.20	0.62	0.26	0.20	−0.06
CPI inflation	−1.16	−1.34	−0.38	−0.04	−2.69	−3.81*	−0.88	−0.90
PCE inflation	−0.37	−1.59	−0.83	−1.03	0.13	−2.54	−0.31	−1.48
IP growth volatility	−0.64	0.11	0.86	0.52	−3.58*	−0.77	−0.06	0.33
CPI inflation volatility	−0.85	−0.63**	−1.40	−1.30	−0.77	−0.42	−0.47	−0.22
PCE inflation volatility	−0.02	−0.33	0.27	0.12	0 .02	0.37	−0.30	−1.14*
Stock return	−1.05	−1.28	−5.24*	−7.02*	0.39	2.53	1.46	1.03
Stock volatility	−1.48*	−0.70	−0.88	−1.27	−2.53	−0.74	−1.34	−0.22
Stock liquidity	−0.60	−0.35	−0.66	−0.58	−1.19	−0.31	−0.21	−0.20*
Default spread	−0.00	−0.25	0.39	0.19	1.25	−0.50	−1.10	−0.45
Term spread	−0.98	−0.83	−2.40	−2.23	−0.67	−0.31	−0.03	−0.02
RF rate difference	−0.20	−0.49	0.21	0.34	−0.78	−2.04	−0.17	−0.22
FF rate difference	−0.14	−0.49	0.23	0.44	−0.97	−1.47	0.06	−0.65
Large	−6.54**	−4.68**	−6.23	−6.81	−15.1**	−10.0*	−1.63	0.28
Mean	0.04	−0.04	−0.15	−0.27	−0.38	−0.17	0.45	0.22
Median	0.08	0.06	0.21	0.25	−0.22	−0.28	0.16	0.12
Trimmed mean	0.10	−0.04	0.11	−0.01	−0.28	−0.36	0.31	0.14
MSFE	0.16	0.08	0.22	0.10	−0.28	−0.10	0.50	0.27

Note: The table reports the increase in out-of-sample R^2 relative to the benchmark when the forecasting model includes economic variables. The numbers in the table are their original values multiplied by 100. The superscripts ***, ** and * indicate that the null of equal predictive ability can be rejected at 1%, 5% and 10%, respectively. "Large" means the model with a large set of forecasting variables. "Roll" and "Recur" represent the rolling and recursive estimation scheme, respectively.

where $\ln\widehat{RV}_{k,t+1}$ is the individual forecast of volatility k and $w_{k,t}$ is the associated combining weight. Following Rapach et al. (2010), forecast combination proceeds with several simple mechanisms.[9] The first two combination schemes are the mean and median across n individual forecasts, respectively. The third approach employs the trimmed mean, which drops the

[9] As indicated in Timmermann (2006), those complicated combining methods generally do not outperform the simple schemes. Thus the subsequent analysis does not include those complicated combined forecasts.

Table 10
Out-of-sample results: ten-year bond return volatility.

	1983–2012		1983–1992		1993–2002		2003–2012	
	Roll	Recur	Roll	Recur	Roll	Recur	Roll	Recur
IP growth	0.86	0.45	0.64	0.76	0.14	−0.58	1.62	0.94
ISM index	0.32	0.67	−0.28	0.30	0.04	0.66	1.09	1.05
Employment growth	0.34	0.03	−0.54	−0.49	0.33	0.27	1.18	0.33
CPI inflation	−1.25	−1.53	−0.02	0.44	−4.08	-3.90^{*}	−0.36	-1.66^{*}
PCE inflation	−1.02	−1.94	−1.50	−1.41	−2.06	−3.28	0.13	-1.52^{*}
IP growth volatility	0.09	−0.09	1.73	1.46	−1.17	−2.02	−0.47	−0.08
CPI inflation volatility	-0.79^{*}	-0.52^{**}	-0.72^{*}	-1.10^{*}	−0.97	−0.29	−0.77	−0.17
PCE inflation volatility	−0.27	−0.19	−0.11	0.12	0.32	0.18	-0.88^{*}	−0.75
Stock return	−0.35	−0.24	−3.25	−3.20	−0.39	1.02	2.31	1.51^{*}
Stock volatility	−0.02	−0.68	0.32	−0.47	−3.20	−2.04	2.00	0.11
Stock liquidity	−0.28	-0.45^{*}	−0.51	−0.55	−0.17	−0.77	−0.17	-0.15^{**}
Default spread	0.78	−0.15	0.24	0.23	1.33	−0.88	0.89	0.03
Term spread	−0.81	−1.04	−2.58	−2.36	−0.26	−0.26	0.35	−0.44
RF rate difference	0.36	−0.65	0.61	0.36	−0.19	-1.88^{*}	0.56	-0.68^{*}
FF rate difference	0.21	-0.84^{*}	−0.22	−0.49	−0.21	−1.18	0.93	-0.94^{**}
Large	−3.40	−3.43	−2.00	−1.46	-12.8^{*}	−12.4	2.02	1.21
Mean	0.57	0.01	0.14	0.12	−0.23	−0.55	1.57^{**}	0.33
Median	0.42^{*}	0.06	0.27	0.37	−0.00	−0.34	0.90^{*}	0.08
Trimmed mean	0.48^{*}	0.02	0.35	0.28	−0.20	−0.53	1.12^{*}	0.19
MSFE	0.65^{*}	0.09	0.36	0.34	−0.15	−0.47	1.64^{**}	0.39^{*}

Note: The table reports the increase in out-of-sample R^2 relative to the benchmark when the forecasting model includes economic variables. The numbers in the table are their original values multiplied by 100. The superscripts ***, ** and * indicate that the null of equal predictive ability can be rejected at 1%, 5% and 10%, respectively. "Large" means the model with a large set of forecasting variables. "Roll" and "Recur" represent the rolling and recursive estimation scheme, respectively.

smallest and the largest forecasts and assigns $w_{k,t}=1/\,(n-2)$ for the remaining members. The final rule follows Stock and Watson (2004) and assigns the combining weights as

$$w_{k,t} = \frac{\phi_{k,t}^{-1}}{\sum_{l=1}^{n}\phi_{l,t}^{-1}},$$

where

$$\phi_{k,t} = \sum_{q=R}^{t-1}\left(\ln RV_{q+1} - \ln\hat{RV}_{k,q+1}\right).$$

This combining procedure assigns more weight on an individual forecast with lower historical MSFE, which means better past performance.

As shown in Tables 8, 9, 10 and 11, some combined forecasts deliver significantly better out-of-sample performance relative to the benchmark. For the longer 1983–2012 period, the superior predictive ability is particularly evident in the volatility forecasts for 1-year Treasuries since all the ΔR_{OS}^2 statistics are positive and several of them are statistically significant. For Treasuries of other maturities, only some volatility forecasts for 10-year Treasuries outperform the benchmark with statistical significance. In other cases, no statistical significance is detected and the ΔR_{OS}^2 statistics turn to be negative in some occasions. Thus the predictive ability of combined forecasts also concentrates on the short end of the term structure. For the other three shorter sample periods, the predictive ability is more evident in some 1- and 10-year volatility forecasts for the relatively turbulent 2003–2012 horizon. Although forecast combination appears to improve the out-of-sample performance, the improvement by forecast combination is still small in terms of economic significance since the largest increase in out-of-sample R^2 is less than 2%. Because combined forecasts more or less inherit the characteristics from their individual members, it is not surprising that the increase in out-of-sample R^2 is not impressive for any combination scheme.

5. Sources of poor forecasting performance

Many empirical applications find that in-sample tests tend to easily reject the null of no predictive ability while out-of-sample tests do not. The forecasting literature has suggested several possible explanations to the lack of out-of-sample predictive ability. First of all, the forecasting ability of a model might not be stable over time. More explicitly, the estimated relationship over one period might not apply to subsequent horizons. This instability of forecast is often caused by structural breaks in the model parameters. Second, in-sample predictive relationship might lack predictive content. Excellent in-sample fit of a forecasting model does not guarantee prominent out-of-sample forecasts. In particular, the in-sample fit of a

Table 11
Out-of-sample results: thirty-year bond return volatility.

	1983–2012		1983–1992		1993–2002		2003–2012	
	Roll	Recur	Roll	Recur	Roll	Recur	Roll	Recur
IP growth	0.82	0.36	0.53	0.87*	0.62	−0.61	1.53	0.68
ISM index	0.56	0.44	0.12	0.16	0.30	0.59	1.27	0.78
Employment growth	0.07	−0.00	−0.79	−0.65	0.18	0.04	0.62	0.41
CPI inflation	−0.96	−1.59*	−1.15	−0.43	−5.22**	−4.84**	0.79	−1.59*
PCE inflation	−1.07	−2.21**	−2.74	−2.77	−4.50*	−5.50**	1.19	−1.36**
IP growth volatility	0.05	−0.32	1.99	1.84	−0.97	−1.49	−0.71	−1.35
CPI inflation volatility	−0.46	−0.36**	−0.27	−0.99*	−1.11	−0.02	−0.50	−0.29
PCE inflation volatility	−0.51	−0.15	−0.52	−0.05	−0.79	−0.32	−0.63	−0.22
Stock return	−0.34	−0.45	−3.92*	−4.37	−0.39	0.65	1.90*	1.39
Stock volatility	1.07	−0.43	0.47	0.13	−2.35	−3.32	3.71	0.41
Stock liquidity	−0.83	−0.59	−2.63	−2.17	−0.82	−0.18	−0.07	−0.05
Default spread	0.49	−0.19	−0.29*	−0.26**	−0.49	−0.10	1.76	−0.29
Term spread	−1.11	−1.17	−3.21	−2.97	−0.79	−0.69	−0.44	−0.83
RF rate difference	0.36	−0.37	0.22	0.22	−0.17	−1.59*	0.89	−0.36
FF rate difference	0.19	−0.59	−0.23	−0.34	−0.21	−0.90	0.76	−0.92*
Large	−2.15	−3.12*	−4.08	−1.64	−12.6**	−13.2**	3.20	−0.77*
Mean	0.54	−0.10	0.01	−0.06	−0.52	−0.51	1.71	0.02
Median	0.45	−0.08	0.18	−0.01	−0.03	0.04	1.11	−0.23
Trimmed mean	0.44	−0.05	0.27	0.11	−0.40	−0.47	1.19	0.02
MSFE	0.64	−0.01	0.35	0.29	−0.41	−0.40	1.82	0.08

Note: The table reports the increase in out-of-sample R^2 relative to the benchmark when the forecasting model includes economic variables. The numbers in the table are their original values multiplied by 100. The superscripts ***, ** and * indicate that the null of equal predictive ability can be rejected at 1%, 5% and 10%, respectively. "Large" means the model with a large set of forecasting variables. "Roll" and "Recur" represent the rolling and recursive estimation scheme, respectively.

predictive regression might be negatively related with its out-of-sample performance. Thus the estimated relationship essentially indicates a wrong direction for the future. The third possibility is overfitting, which refers to the inclusion of irrelevant explanatory variables in a forecasting model. These nuisance regressors might improve the fit of a predictive regression while deteriorate its out-of-sample performance.

To explore the sources of poor bond return volatility forecasts, I apply the methodology proposed by Rossi and Sekhposyan (2011) to decompose the relative out-of-sample performance into components that measures forecast instability, predictive content and overfitting. The authors also develop the procedures to test the significance for these three asymptotically uncorrelated components.[10] First of all, the forecast instability component is defined as

$$A_{\tau,P} = \frac{1}{m}\sum_{t=R+\tau-m}^{R+\tau-1}\hat{L}_{t+1} - \frac{1}{P}\sum_{t=R}^{T}\hat{L}_{t+1}, \tag{7}$$

where R is the length of estimation period, P is the length of the forecasting period, $T=R+P-1$ is the length of the whole sample period, $\tau=m,m+1,\ldots,P$ is the length of a subset of the forecasting period, and $\hat{L}_{t+1}$ is the squared forecast error difference between the benchmark and the augmented model. This component measures the variation of out-of-sample performance in a given period over time. According to this definition, $E\,(A_{\tau,P})$ deviates from zero if the out-of-sample forecasting ability considerably changes over time. Thus the rejection of $E\,(A_{\tau,P})=0$ means substantial instability in the out-of-sample forecasts. To derive the other two components, first run the regression

$$\hat{L}_{t+1} = \gamma\hat{\mathcal{L}}_t + u_{t+1}, \tag{8}$$

where $\hat{\mathcal{L}}_t$ measures in-sample fit and u_{t+1} is the residual. The predictive content component is defined as

$$B_P = \left(\frac{1}{P}\sum_{t=R}^{T}\hat{\mathcal{L}}_t\right)\hat{\gamma}, \tag{9}$$

where $\hat{\gamma}$ is the estimates of γ. This component indicates whether in-sample fit contains sufficient and correct information for out-of-sample forecasting performance. If B_P is close to zero, the in-sample result lacks predictive content since it is nearly uncorrelated with out-of-sample performance. If B_P is significantly different from zero, the predictive content depends on

[10] According to the authors, the test statistic for $E\,(B_P)=0$ or $E\,(U_P)=0$ is asymptotically $N(0, 1)$. When testing $E\,(A_{\tau,P})=0$, the asymptotical distribution is non-standard. The critical values are provided by the authors.

Table 12
Forecast decomposition: rolling scheme.

	1-Year			5-Year			10-Year			30-Year		
	$A_{\tau,P}$	B_P	U_P	$A_{\tau,P}$	B_P	U_P	$A_{\tau,P}$	B_P	U_P	$A_{\tau,P}$	B_P	U_P
IP growth	0.06	0.84	−0.89	−0.32	0.10	1.64	−0.03	0.54	1.28	0.03	0.79	1.27
ISM index	−0.32	0.01	−0.80	−0.24	−0.56	−0.16	−0.16	−0.83	0.66	−0.11	0.28	0.94
Employment growth	0.01	0.53	−1.01	0.22	−0.97	0.64	0.04	0.43	0.60	0.37	1.42	−0.12
CPI inflation	0.16	0.46	−0.36	−0.13	−1.59	−1.24	−0.55	−1.48	−1.04	−0.57	−1.58	−0.86
PCE inflation	0.43	0.09	0.14	0.01	1.27	−0.35	−0.10	−1.32	−0.84	−0.33	−1.74*	−0.95
IP growth volatility	−0.90	1.84*	1.03	−0.77	−1.24	−0.96	−0.38	−1.47	0.26	−0.32	−0.98	0.19
CPI inflation volatility	−0.27	0.42	0.75	0.71	−0.28	−1.64	0.71	0.81	−1.70*	0.46	2.20**	−1.48
PCE inflation volatility	0.37	−0.34	−0.70	−0.41	−0.48	−0.01	0.03	−0.22	−0.69	0.04	1.34	−1.62
Stock return	0.99	−0.32	−1.03	0.62	0.59	−1.01	−0.01	−0.07	−0.41	−0.03	−0.26	−0.46
Stock volatility	0.15	−1.01	−0.92	0.30	0.25	−1.92*	−0.53	1.03	−0.14	−0.45	1.03	0.43
Stock liquidity	−0.02	0.89	−0.53	−0.20	0.96	−1.20	0.12	−0.18	−0.66	0.05	−0.56	−0.97
Default spread	−1.29	−0.41	2.94***	0.85	0.56	−0.06	0.09	−0.53	1.11	−0.20	1.31	0.12
Term spread	−0.36	−0.81	−0.72	0.46	−1.42	−0.97	0.52	−1.35	−0.83	0.69	−1.02	−1.25
RF rate difference	−0.64	0.46	1.55	−0.21	−0.36	−0.38	0.10	1.05	0.60	0.12	1.41	0.41
FF rate difference	−0.13	−0.07	0.78	−0.07	0.54	−0.28	0.04	1.24	0.38	0.22	1.71*	0.23
Large	0.16	−0.73	−0.59	−0.01	0.43	−2.44***	−0.42	−0.10	−1.39	−0.47	−0.81	−1.04

Note: The table reports the test statistics for the hypothesis $E\,(A_{\tau,P})=0$, $E\,(B_P)=0$ and $E\,(U_P)=0$, respectively. Out-of-sample forecasts are generated by rolling estimation scheme. The time variation of forecasting performance is measured by averaging the relative predictive ability over 240 months. "Large" means the model with a large set of forecasting variables. The superscripts ***, ** and * indicate that the null of equal predictive ability can be rejected at 1%, 5% and 10%, respectively.

the sign of B_P and the average out-of-sample performance $(1/P)\sum_{t=R}^{T}\hat{L}_{t+1}$. When B_P has the same sign as $(1/P)\sum_{t=R}^{T}\hat{L}_{t+1}$, in-sample and out-of-sample outcomes are consistent. Good (bad) in-sample fit also correctly indicates superior (inferior) out-of-sample predictive ability. When B_P has the opposite sign with $(1/P)\sum_{t=R}^{T}\hat{L}_{t+1}$, the predictive content is misleading since in-sample fit is negatively related with out-of-sample performance. Finally, the overfitting component is defined as

$$U_P = \frac{1}{P}\sum_{t=R}^{T} u_{t+1}. \tag{10}$$

This component captures the part of out-of-sample performance that is not explained by in-sample fit. The rejection of $E\,(U_P)=0$ suggests that the out-of-sample forecasting ability of a model is not reflected by in-sample fit. Note that Eqs. (8), (9) and (10) imply that

$$\frac{1}{P}\sum_{t=R}^{T}\hat{L}_{t+1} = B_P + U_P. \tag{11}$$

Thus the sum of predictive content and overfitting components is exactly the average forecasting performance.

Tables 12 and 13 report the test statistics for the significance of each component. The results indicate that the inferior bond return volatility forecasts are mostly caused by overfitting from two reasons. First, the hypothesis $E\,(U_P)=0$ is rejected for every significantly inferior forecast while the other two components are almost silent. Second, these rejections mean that $E\,(U_P)$ is significantly negative for these forecasts. Because the average out-of-sample performance is exactly the sum of B_P and U_P, the forecasting performance of these models must be significantly worse than the benchmark if B_P is essentially zero. For those volatility forecasts featuring negative U_P, including economic variables in these predictive regressions enhances in-sample fit but penalizes their out-of-sample predictive ability. On the other hand, the misleading predictive content explains the poor forecasts of 30-year bond return volatility predicted by the variation of CPI inflation under the recursive scheme. In this case, the both negative $E\,(B_P)$ and $(1/P)\sum_{t=R}^{T}\hat{L}_{t+1}$ means that the deterioration of in-sample fit correctly predicts inferior out-of-sample performance. Otherwise, this component does not explain other poor volatility forecasts since the null of $E\,(B_P)=0$ cannot be rejected. Although the in-sample evidence implies time-varying forecasting ability of these economic variables, the hypothesis $E\,(A_{\tau,P})=0$ cannot be rejected in any case. Thus forecast instability is not evident enough to be responsible for poor bond return volatility forecasts.[11] In practice, this diagnosis suggests not to include a large number of economic variables in the bond return volatility forecasting model, particularly when making out-of-sample volatility forecasts of intermediate and long maturities.

Based on the above diagnosis, it is natural to investigate whether reducing the dimension of forecasting variables

[11] Although this conclusion is based on $\tau=240$, it is generally invariant to different values of τ. The hypothesis $E\,(A_{\tau,P})=0$ cannot be rejected in almost every case even if $\tau=60$ or 120.

Table 13
Forecast decomposition: recursive scheme.

	1-Year			5-Year			10-Year			30-Year		
	$A_{\tau,P}$	B_P	U_P	$A_{\tau,P}$	B_P	U_P	$A_{\tau,P}$	B_P	U_P	$A_{\tau,P}$	B_P	U_P
IP growth	0.57	0.47	−0.59	−0.21	−1.39	1.48	−0.17	−0.72	0.71	−0.30	0.56	1.26
ISM index	−0.16	−0.22	0.00	−0.37	−1.18	−1.00	−0.29	−1.28	1.33	−0.48	−0.29	1.45
Employment growth	−0.13	−0.02	−0.82	0.25	−0.47	0.31	−0.14	−0.04	0.09	−0.30	1.10	−0.07
CPI inflation	0.46	−1.22	−0.83	−0.18	−2.33***	−1.16	0.02	−2.18**	−1.08	0.03	−1.32	−1.72*
PCE inflation	0.21	−0.84	−0.29	−0.14	−2.06**	−1.12	−0.12	−2.49***	−1.32	−0.20	−1.63	−2.25**
IP growth volatility	−0.78	1.55	1.53	−0.21	−0.04	0.41	0.10	−0.24	−0.09	0.38	0.22	−0.37
CPI inflation volatility	−1.55	0.53	1.43	−0.26	−0.16	−2.24**	−0.25	−0.80	−2.40***	0 .14	−1.71*	−2.21**
PCE inflation volatility	0.82	0.55	−0.69	0.61	0.36	−0.72	0.54	−0.74	−0.47	0.30	0.01	−1.11
Stock return	0.13	−0.86	−0.51	−0.47	0.83	−0.94	−0.69	−0.93	−0.23	−0.79	−0.84	−0.55
Stock volatility	−0.24	−1.04	−0.98	−0.10	−0.77	−0.91	−0.25	0.84	−1.16	−0.41	0.75	−0.83
Stock liquidity	−0.25	0.62	−1.83*	−0.10	0.80	−1.24	−0.25	0.54	−1.71*	−0.22	0.92	−0.96
Default spread	−0.82	0.36	1.19	0.33	1.04	−0.68	−0.08	−0.26	−0.58	0.34	0.27	−1.61
Term spread	−0.71	−0.01	0.47	−0.24	−1.52	−0.89	−0.11	−1.52	−0.92	0.00	−1.11	−1.27
RF rate difference	−0.93	−0.76	2.03**	−0.17	0.26	−0.85	0.01	−0.62	−1.38	0.06	−0.56	−1.40
FF rate difference	−0.57	−0.37	1.53	0.07	−0.25	−0.85	0.10	−0.72	−1.70*	0.33	−0.88	−1.53
Large	0.09	−0.75	−1.19	−0.52	0.19	−1.97**	−0.42	0.90	−1.45	−0.33	0.09	−1.95*

Note: The table reports the test statistics for the hypothesis $E\ (A_{\tau,P})=0$, $E\ (B_P)=0$ and $E\ (U_P)=0$, respectively. Out-of-sample forecasts are generated by recursive estimation scheme. The time variation of forecasting performance is measured by averaging the relative predictive ability over 240 months. "Large" means the model with a large set of forecasting variables. The superscripts ***, ** and * indicate that the null of equal predictive ability can be rejected at 1%, 5% and 10%, respectively.

improves the forecasting performance. The evaluation proceeds with extracting several principal components from these economic variables. Table 14 reports the out-of-sample performance of the first two (PC1, PC2) and the fifth (PC5) principal components. This is because PC1 and PC2 explain the most of variation in these economic variables and PC5 Granger causes bond return volatility of all maturities.[12] Nevertheless, these principal components simply increase the out-of-sample R^2 by a few percent relative to the benchmark and the null of equal predictive ability cannot be rejected in most cases. It is also noted that the inclusion of PC2 sometimes leads to significantly worse volatility forecasts. This poor performance implies that the predictive relationship between inflation and bond return volatility in the estimation period (1963–1982) is not applied to the period of Great Moderation. In brief, variable reduction by extracting some principal components does not lead to better out-of-sample bond return volatility forecasts.

Readers might wonder why forecast instability explains little of the difference between in-sample and out-of-sample forecasts even if the forecasting ability of economic variables varies across different historical periods. Although this conclusion comes a formal statistical test, it is possible to get the intuition from the out-of-sample performance under different estimation schemes. If forecast instability were evident, the predictive relationship based on rolling samples would generate significantly superior volatility forecasts since recursive samples include more outdated information that might not be appropriate for the future. However, the out-of-sample results (see Tables 8-11) do not suggest a remarkable difference between forecasts based on rolling and recursive samples. Sometimes one scheme outperforms the other, and the null of no superior predictive ability cannot be rejected for most forecasting models. Although the forecasting ability of economic variables does fluctuates over time, this variation might not be enough to explain the difference between in-sample fit and out-of-sample prediction.

6. Conclusion

Using the CRSP Fixed Term Indices data for the 1963–2012 period, this study investigates whether including various economic variables in predictive regressions improves the forecasts for monthly US bond return volatility. Among various forecasting variables, stock return or movements in the Federal Funds rate unambiguously Granger causes bond return volatility of various maturities. Variables of inflation, credit condition or the term structure also Granger causes short-term bond return volatility, while measures of output growth, productivity or employment condition sometimes Granger causes volatility of intermediate and long maturities. The forecasting ability is particularly evident at the short end of the term structure or in the relatively turbulent historical periods. On the other hand, the economic significance is a question since the improvement of in-sample fit is simply a few percent. The evidence of out-of-sample predictive ability is generally weaker since only a few forecasts significantly beat the benchmark, but several forecast combination schemes sometimes improve the out-of-sample performance with modest economic gain. To analyze the difference between in-sample and out-

[12] The in-sample results using these principal components are available upon request. Their estimated coefficients are also mostly significant on the short end of the term structure.

Table 14
Out-of-sample results: principal components.

	1983–2012		1983–1992		1993–2002		2003–2012	
	Roll	Recur	Roll	Recur	Roll	Recur	Roll	Recur
1-Year								
PC1	0.05	0.95	0.09	0.05	−0.90	−1.19*	0.52	2.90**
PC2	−0.58	−1.47	−4.67	−7.01	1.49	2.59	−0.83	−2.88**
PC5	0.00	0.07	4.43	5.56**	−2.11	−1.44	−0.23	−0.71
All	0.08	−0.22	1.40	0.24	1.19	2.82	−0.77	−1.95
5-Year								
PC1	−0.60	−0.38	−0.36	−0.23	−0.78	−1.09	−0.69	−0.06
PC2	−1.45	−2.89*	−3.97	−4.61	−0.19	−2.51	−0.18	−1.76
PC5	0.55	0.87	4.24	4.09*	−5.03	−3.80	1.00	1.13
All	0.07	−0.97	2.45	1.30	−2.81	−4.75	−0.10	−0.50
10-Year								
PC1	0.87	−0.32	0.25	0.16	0.77	−0.97	1.57	−0.29**
PC2	−1.19	−2.34*	−2.85	−3.14	−0.48	−2.37	−0.25	−1.72
PC5	0.61	−0.37	2.63	2.04	−1.61	−4.46	0.42	0.43
All	1.25	−1.72	1.58	0.81	0.11	−5.71	1.85	−1.19
30-Year								
PC1	1.06	−0.42	0.43	0.44	0.42	−1.14	2.34	−0.86*
PC2	−1.80**	−2.59**	−4.41	−4.72	−3.24*	−4.29*	−0.27	−1.68**
PC5	0.33	−0.08	2.89	2.81	−1.26	−3.38	−0.42	−0.41
All	0.41	−2.21**	−0.06	−0.34	−2.17	−6.52**	2.21	−2.49

Note: The table reports the increase in out-of-sample R^2 by the forecasting model with the first, second and fifth principal components (PC1, PC2 and PC5) of the economic variables. The numbers in the table are their original values multiplied by 100. The superscripts ***, ** and * indicate that the null of equal predictive ability can be rejected at 1%, 5% and 10%, respectively. “All” means that PC1, PC2 and PC5 are all included as predictors. “Roll” and “Recur” represent the rolling and recursive estimation scheme, respectively.

of-sample predictability, this paper employs the decomposition of forecasts in the manner of Rossi and Sekhposyan (2011). The associated statistical tests indicate that overfitting is primarily responsible for this difference, and variable reduction by extracting principal components does not solve this problem.

According to the above analysis, some economic variables are conducive to bond return volatility prediction, but these augmented models generally do not outperform the benchmark model and the forecasting improvements are not impressive. This conclusion raises a question of the link between bond markets and macroeconomic fundamentals. Does the weak forecasting ability characterize limited effects of macroeconomic variables on bond markets, or the link is actually essential but hidden behind the persistence of bond return volatility? Based on the evidence of countercyclical bond return volatility, the hypothesis of hidden prediction information appears to be preferable. If the predictive relationship is simply obscured by the persistence, it would be important to disentangle the forecasting information in the economic variables from this time dependency in the future research.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens (2001). The distribution of realized stock volatility. *Journal of Financial Economics*, *61*, 43–76.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, *96*, 42–55.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys (2003). Modelling and forecasting realized volatility. *Econometrica*, *71*, 579–625.
Ang, A., Piazzesi, M., & Wei (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, *131*, 359–403.
Bansal, R., & Shaliastovich (2013). A long-run risks explanation of predictability puzzles in bond and currency markets. *Review of Financial Studies*, *26*, 1–33.
Bollerslev, T., Caic, J., & Song (2000). Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the US treasury bond market. *Journal of Empirical Finance*, *7*, 37–55.
Brenner, M., Pasquariello, P., & Subrahmanyam (2009). On the volatility and comovement of U.S. financial markets around macroeconomic news announcements. *Journal of Financial and Quantitative Analysis*, *44*, 1265–1289.
Brunnermeier, M. K., & Pedersen (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, *22*, 2201–2238.
Campbell, J. Y., & Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*, 1509–1531.
Christiansen, C., Schmeling, M., & Schrimpf (2012). A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics*, *27*, 956–977.
Fleming, J., Kirby, C., & Ostdiek (1998). Information and volatility linkages in the stock, bond, and money markets. *Journal of Financial Economics*, *49*, 111–137 .
Fontaine, J.-S., & Garcia (2012). Bond liquidity premia. *Review of Financial Studies*, *25*, 1207–1254.
Giacomini, R., & White (2006). Tests of conditional predictive ability. *Econometrica*, *74*, 1545–1578.

Goyal, A., & Welch (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*, 1455–1508.
Goyenko, R. Y., & Ukhov (2009). Stock and bond market liquidity: A long-run empirical analysis. *Journal of Business and Economic Statistics*, *12*, 167–176.
Jones, C. M., Lamont, O., & Lumsdaine (1998). Macroeconomic news and bond market volatility. *Journal of Financial Economics*, *47*, 315–337.
Ludvigson, S. C., & Ng (2009). Macro factors in bond risk premia. *Review of Financial Studies*, *22*, 5027–5067.
Mueller, P., Vedolin, A., & Yen (2011). Bond variance risk premia. In *Working paper.* London School of Economics: London School of Economics, 2011.
Pástor, L., & Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*, 642–685.
Paye, B. S. (2012). 'Déjà Vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, *106*, 527–546.
Rapach, D. E., Strauss, J. K., & Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, *23*, 821–862.
Rossi, B., & Sekhposyan (2011). Understanding models' forecasting performance. *Journal of Econometrics*, *164*, 158–172.
Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance*, *44*, 1115–1153.
Stock, J. H., & Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*, 405–430.
Timmermann, A. (2006). Forecast combinations. In *Handbook of economic forecasting*. Vol. 1.
Viceira, L. M. (2012). Bond risk, bond return volatility and the term structure of interest rates. *International Journal of Forecasting*, *28*, 97–117.