Bond Return Predictability: Economic Value and Links to the Macroeconomy*

Antonio Gargano†
University of Melbourne

Davide Pettenuzzo‡
Brandeis University

Allan Timmermann§
University of California San Diego

July 23, 2014

Abstract

Studies of bond return predictability find a puzzling disparity between strong statistical evidence of return predictability and the failure to convert return forecasts into economic gains. We show that resolving this puzzle requires accounting for important features of bond return models such as time varying parameters and volatility dynamics. A three-factor model comprising the Fama and Bliss (1987) forward spread, the Cochrane and Piazzesi (2005) combination of forward rates and the Ludvigson and Ng (2009) macro factor generates notable gains in out-of-sample forecast accuracy compared with a model based on the expectations hypothesis. Importantly, we find that such gains in predictive accuracy translate into higher risk-adjusted portfolio returns after accounting for estimation error and model uncertainty, as evidenced by the performance of model combinations. Finally, we find that bond excess returns are predicted to be significantly higher during periods with high inflation uncertainty and low economic growth and that the degree of predictability rises during recessions.

JEL codes: G11, G12, G17

1 Introduction

Treasury bonds play an important role in many investors' portfolios so an understanding of the risk and return dynamics for this asset class is of central economic importance.[1] Some studies document significant in-sample predictability of Treasury bond excess returns for 2-5 year

*We thank Blake LeBaron and seminar participants at USC, University of Michigan, Central Bank of Belgium, ESSEC Paris and Econometric Society Australasian Meeting (ESAM) for comments on the paper.

†University of Melbourne, Building 110, Room 11.042, 198 Berkeley Street, Melbourne, 3010. Email: antonio.gargano@unimelb.edu.au

‡Brandeis University, Sachar International Center, 415 South St, Waltham, MA, Tel: (781) 736-2834. Email: dpettenu@brandeis.edu

§University of California, San Diego, 9500 Gilman Drive, MC 0553, La Jolla CA 92093. Tel: (858) 534-0894. Email: atimmerm@ucsd.edu.

[1] According to the Securities Industry and Financial Markets Association, the size of the U.S. Treasury bond market was $11.9 trillion in 2013Q4. This is almost 30% of the entire U.S. bond market which includes corporate debt, mortgage and municipal bonds, money market instruments, agency and asset-backed securities.

maturities by means of variables such as forward spreads (Fama and Bliss (1987)), yield spreads (Campbell and Shiller (1991)), a linear combination of forward rates (Cochrane and Piazzesi (2005)) and factors extracted from a cross-section of macroeconomic variables (Ludvigson and Ng (2009)).

While empirical studies suggest that there is strong statistical evidence in support of bond return predictability, there is so far little evidence that such predictability could have been used in real time to improve investors' economic utility. Notably, Thornton and Valente (2012) find that forward spread predictors, when used to guide the investment decisions of an investor with mean-variance preferences, do not lead to higher out-of-sample Sharpe ratios or higher economic utility compared with decisions based on a no-predictability expectations hypothesis (EH) model. Sarno et al. (2014) reach a similar conclusion.

To address this puzzling contradiction between the statistical and economic evidence on bond return predictability, we propose a new empirical modeling approach that generalizes the existing literature in economically insightful ways. Modeling bond return dynamics requires adding several features that are absent from the regression models used in the existing literature. First, bond prices, and thus bond returns, are sensitive to monetary policy and inflation prospects, both of which are known to shift over time.[2] This suggests that it is important to adopt a framework that accounts for time varying parameters and even for the possibility that the forecasting model may shift over time, requiring that we allow for model uncertainty. Second, uncertainty about inflation prospects changes over time and the volatility of bond yields has also undergone shifts–most notably during the Fed's monetarist experiment from 1979-1982–underscoring the need to allow for time-varying volatility.[3] Third, risk-averse bond investors are concerned not only with the most likely outcomes but also with the degree of uncertainty surrounding future bond returns, indicating the need to model the full probability distribution of bond returns.

The literature on bond return predictability has noted the importance of parameter estimation error, model instability, and model uncertainty. However, no study on bond return predictability has so far addressed how these considerations, jointly, impact the results. To accomplish this, we propose a novel Bayesian approach that brings several advantages to inference about the return prediction models and to their use in portfolio allocation analysis.

Our approach allows us, first, to integrate out uncertainty about the unknown parameters and to evaluate the effect of estimation error on the results. Estimation errors turn out to be important for understanding our results. For example, the improved performance associated with more flexible specifications such as time varying parameter models sometimes comes at the

[2] Stock and Watson (1999) and Cogley and Sargent (2002) find strong evidence of time-variations in a Phillips curve model for U.S. inflation.

[3] Sims and Zha (2006) and Cogley et al. (2010) find that it is important to account for time-varying volatility when modeling U.S. macroeconomic dynamics.

cost of larger estimation errors. Conversely, models with time varying volatility tend to have more precisely estimated parameters as they reduce the weight on periods with highly volatile, and thus noisy, bond returns.[4]

Second, our approach produces predictive densities of bond excess returns. This allows us to analyze the economic value of bond return predictability from the perspective of an investor with power utility. Thornton and Valente (2012) are limited to considering mean-variance utility since they only model the first two moments of bond returns.[5]

Third, we allow for time-varying volatility in the bond excess return model. Bond market volatility spiked during the monetarist experiment from 1979 to 1982, but we find clear advantages from allowing for stochastic volatility beyond this episode, particularly for bonds with shorter maturities.

A fourth advantage of our approach is that it allows for time-variation in the regression parameters. Thornton and Valente (2012) (p. 3157) report that their results are "indicative of a considerable time variation in the parameter estimates." Our results concur with this and we find that the slope coefficients on both the yield spreads and the macrofactors vary considerably during our sample.

Fifth, we address model uncertainty through model combination methods. We consider equal-weighted averages of predictive densities, Bayesian model averaging, as well as combinations based on the optimal pooling method of Geweke and Amisano (2011). The latter forms a portfolio of the individual prediction models using weights that reflect the models' posterior probabilities. Models that are more strongly supported by the data get a larger weight in this average. The model combination results are better than the results for the individual models and thus suggest that model uncertainty can be effectively addressed through combination methods.

As emphasized by Johannes et al. (2014), an ensemble of such extensions to the constant mean, constant volatility model is required to establish evidence of significant out-of-sample return predictability. For example, accounting for parameter estimation error is no guarantee for good out-of-sample results. Model uncertainty also plays an important role as forecasting performance varies considerably across different prediction models. Moreover, we find that the importance of the enhancements varies with the maturity of the underlying bonds: volatility dynamics are particularly important for the short (2-3 year) maturities, while time varying parameters are more important for bonds with longer (4-5 year) maturities.

Our empirical analysis uses the daily treasury yield data from Gurkaynak et al. (2007) to

[4] Altavilla et al. (2014) find that an exponential tilting approach helps improve the accuracy of out-of-sample forecasts of bond yields. While their approach is not Bayesian, their tilting approach also attenuates the effect of estimation error on the model estimates.

[5] Sarno et al. (2014) use an approximate solution to compute optimal portfolio weights under power utility. They do not find evidence of economically exploitable return predictability but also do not consider parameter uncertainty.

construct monthly excess returns for bond maturities between two and five years over the period 1962-2011. While previous studies have focused on the annual holding period, we find that focusing on the higher frequency affords several advantages. Most obviously, it considerably expands the number of non-overlapping observations, a point of considerable importance given the importance of parameter estimation errors. Moreover, it allows us to identify short-lived dynamics in both first and second moments of bond returns which could be missed by models of annual returns. We find this to be an important consideration, particularly around the time of the financial crisis of 2008 during which bond market returns became quite volatile and around turning points of the economic cycle.

We conduct our analysis in the context of a three-variable model that unifies studies in the existing literature. Specifically, this model includes the Fama-Bliss forward spread, the Cochrane-Piazzesi linear combination of forward rates, and a macro factor constructed using the methodology of Ludvigson and Ng (2009). Each variable is weighted according to its ability to improve on the predictive power of the bond return equation. Since forecasting studies have found that simpler models often do well in out-of-sample experiments, we also consider simpler univariate and bivariate models that include one or two predictors.[6]

To assess the statistical evidence on bond return predictability, we use our models to generate out-of-sample forecasts over the period 1990-2011. Our return forecasts are based on recursively updated parameter estimates and use only historically available information, thus allowing us to assess how valuable the model forecasts would have been to investors in real time. Compared to the benchmark EH model that assumes no return predictability, consistent with Ludvigson and Ng (2009) we find that many of the return predictability models generate significantly positive out-of-sample R^2 values.[7] Interestingly, the Bayesian return prediction models generally perform better than the least squares counterparts so far explored in the literature.

Turning to the economic value of such out-of-sample forecasts, we next consider the portfolio choice between a risk-free Treasury bill versus a bond with 2-5 years maturity for an investor with power utility. We find that the best return prediction models that account for volatility dynamics and changing parameters deliver sizeable gains in certainty equivalent returns relative to an EH model that assumes no predictability of bond returns, particularly in the absence of tight constraints on the portfolio weights.

These findings allow us to reconcile the statistical and economic evidence of bond return predictability. There are several reasons why our findings differ from studies such as Thornton

[6] Other studies considering macroeconomic determinants of the term structure of interest rates include Ang and Piazzesi (2003), Ang et al. (2007), Bikbov and Chernov (2010), Dewachter et al. (2014), Duffee (2011) and Joslin et al. (2014).

[7] Our evaluation uses the out-of-sample R^2 measure proposed by Campbell and Thompson (2008) that compares the sum of squared forecast errors to those from the EH model that includes only an (recursively estimated) intercept term.

and Valente (2012) and Sarno et al. (2014) which argue that the statistical evidence on bond return predictability does not translate into economic gains. Allowing for stochastic volatility and time varying parameters, while accounting for parameter estimation error, leads to important gains in economic performance for many models.[8] Our results on forecast combinations also suggest the importance of accounting for model uncertainty and changes in which prediction model performs best at a given point in time.

To interpret the economic sources of our findings on bond return predictability, we analyze the extent to which such predictability is concentrated in certain economic states and whether it is correlated with variables we would expect to be key drivers of time varying bond risk premia. We find strong evidence that bond return predictability is stronger in recessions than during expansions, consistent with similar findings for stock returns by Henkel et al. (2011) and Dangl and Halling (2012). Economic theory suggests that treasury bond risk premia should be driven by time-varying inflation uncertainty as well as variations in the market price of this source of risk. Using data from survey expectations we find that our bond excess return forecasts are strongly negatively correlated with economic growth prospects (thus being higher during recessions) and strongly positively correlated with inflation uncertainty. This suggests that our bond return forecasts are, at least in part, driven by time-varying risk premia.

The outline of the paper is as follows. Section 2 describes the construction of the bond data, including bond returns, forward rates and the predictor variables. Section 3 sets up the prediction models and introduces our Bayesian estimation approach. Section 4 presents both full-sample and out-of-sample empirical results on bond return predictability. Section 5 assesses the economic value of bond return predictability for a risk averse investor when this investor uses the bond return predictions to form a portfolio of risky bonds and a risk-free asset. This section also analyzes economic sources of bond return predictability such as recession risk and time variations in inflation uncertainty. Section 6 presents model combination results and Section 7 concludes.

2 Data

This section describes how we construct our monthly series of bond returns and introduces the predictor variables used in the bond return models.

2.1 Returns and Forward Rates

Previous studies on bond return predictability such as Cochrane and Piazzesi (2005), Ludvigson and Ng (2009) and Thornton and Valente (2012) use overlapping 12-month returns data.

[8] Thornton and Valente (2012) use a rolling window to update their parameter estimates but do not have a formal model that predicts future volatility or parameter values.

This overlap induces strong serial correlation in the regression residuals. To handle this issue, we reconstruct the yield curve at the daily frequency starting from the parameters estimated by Gurkaynak et al. (2007), who rely on methods developed in Nelson and Siegel (1987) and Svensson (1994). Specifically, the time t zero coupon log yield on a bond maturing in n years, $y_t^{(n)}$, gets computed as[9]

$$\begin{aligned} y_t^{(n)} &= \beta_0 + \beta_1 \frac{1-\exp\left(-\frac{n}{\tau_1}\right)}{\frac{n}{\tau_1}} + \beta_2 \left[\frac{1-\exp\left(-\frac{n}{\tau_1}\right)}{\frac{n}{\tau_1}} - \exp\left(-\frac{n}{\tau_1}\right)\right] \\ &\quad + \beta_3 \left[\frac{1-\exp\left(-\frac{n}{\tau_2}\right)}{\frac{n}{\tau_2}} - \exp\left(-\frac{n}{\tau_2}\right)\right]. \end{aligned} \tag{1}$$

The parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2)$ are provided by Gurkaynak et al. (2007), who report daily estimates of the yield curve from June 1961 onward for the entire maturity range spanned by outstanding Treasury securities. We consider maturities ranging from 12 to 60 months and, in what follows, focus on the last day of each month's estimated log yields.[10]

Denote the frequency at which returns are computed by h, so $h = 1, 3$ for the monthly and quarterly frequencies, respectively. Also, let n be the bond maturity in years. For $n > h/12$ we compute returns and excess returns, relative to the $h-$period T-bill rate[11]

$$r_{t+h/12}^{(n)} = p_{t+h/12}^{(n-h/12)} - p_t^{(n)} = n y_t^{(n)} - (n - h/12) y_{t+h/12}^{(n-h/12)}, \tag{2}$$

$$rx_{t+h/12}^{(n)} = r_{t+h/12}^{(n)} - y_t^{h/12}(h/12). \tag{3}$$

Similarly, forward rates are computed as[12]

$$f_t^{(n-h/12,n)} = p_t^{(n-h/12)} - p_t^{(n)} = n y_t^{(n)} - (n - h/12) y_t^{(n-h/12)}. \tag{4}$$

2.2 Data Summary

Our bond excess return data span the period from 1962:01 through 2011:12. We focus our analysis on the monthly holding period which offers several advantages over the annual returns data which have been the focus of most studies in the literature on bond return predictability. Most obviously, using monthly rather than annual data provides a sizeable increase in the number of data points available for model estimation. This is important in light of the low

[9] The third term was excluded from the calculations prior to January 1, 1980.

[10] The data is available at http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html. Because of idiosyncrasies at the very short end of the yield curve, we do not compute yields for maturities less than twelve months. For estimation purposes, the Gurkaynak et al. (2007) curve drops all bills and coupon bearing securities with a remaining time to maturity less than 6 months, while downweighting securities that are close to this window.

[11] The formulas assume that the yields have been annualized, so we multiply $y_t^{(h/12)}$ by $h/12$.

[12] For $n = h/12$, $f_t^{(n,n)} = n y_t^{(n)}$ and $y_t^{(n-h/12)} = y_t^{(0)}$ equals zero because $P_t^{(0)} = 1$ and its logarithm is zero.

power of the return prediction models. Second, some of the most dramatic swings in bond prices occur over short periods of time lasting less than a year–e.g., the effect of the bankruptcy of Lehman Brothers on September 15, 2008–and are easily missed by models focusing on the annual holding period. This point is also important for the analysis of how return predictability is linked to recessions versus expansions; bond returns recorded at the annual horizon easily overlook important variations around turning points of the economic cycle.

Figure 1 plots monthly bond returns for the 2, 3, 4, and 5-year maturities, computed in excess of the 1-month T-bill rate. All four series are notably more volatile during 1979-82 and the volatility clearly increases with the maturity of the bonds. Table 1 presents summary statistics for the four monthly excess return series. Returns on the shortest maturities are right-skewed and fat-tailed, more so than the longer maturities. This observation suggests that it is inappropriate to use models that assume a normal distribution for bond returns.

2.3 Predictor variables

Our empirical strategy entails regressing bond excess returns on a range of the most prominent predictors proposed in the literature on bond return predictability. Specifically, we consider forward spreads as proposed by Fama and Bliss (1987), a linear combination of forward rates as proposed by Cochrane and Piazzesi (2005), and a linear combination of macro factors, as proposed by Ludvigson and Ng (2009). We briefly explain how we construct these factors.

The Fama-Bliss (FB) forward spreads are computed as

$$fs_t^{(n,h)} = f_t^{(n-h/12,n)} - y_t^{(h/12)}(h/12). \tag{5}$$

The Cochrane-Piazzesi (CP) factor is given as a linear combination of forward rates computed as

$$CP_t^h = \hat{\boldsymbol{\gamma}}^{h\prime}\mathbf{f}_t^{(\mathbf{n}-h/12,\mathbf{n})}, \tag{6}$$

where

$$\mathbf{f}_t^{(\mathbf{n}-h/12,\mathbf{n})} = \left[f_t^{(n_1-h/12,n_1)}, f_t^{(n_2-h/12,n_2)}, ..., f_t^{(n_k-h/12,n_k)}\right].$$

Here $\mathbf{n} = [1, 2, 3, 4, 5]$ denotes the vector of maturities measured in years. As in Cochrane and Piazzesi (2005), the coefficient vector $\hat{\boldsymbol{\gamma}}$ is estimated from

$$\frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} = \gamma_0^h + \gamma_1^h f_t^{(1-1/12,1)} + \gamma_2^h f_t^{(2-1/12,2)} + \gamma_3^h f_t^{(3-1/12,3)} + \gamma_4^h f_t^{(4-1/12,4)} + \gamma_5^h f_t^{(5-1/12,5)} + \overline{\varepsilon}_{t+h/12}. \tag{7}$$

Ludvigson and Ng (2009) propose to use macro factors to predict bond returns. Suppose we observe a $T \times M$ panel of macroeconomic variables $\{x_{i,t}\}$ generated by a factor model

$$x_{i,t} = \kappa_i g_t + \epsilon_{i,t}, \tag{8}$$

where g_t is an $s \times 1$ vector of common factors and $s << M$. The unobserved common factor, g_t is replaced by an estimate, $\hat{g}_t$, obtained using principal components analysis. Following Ludvigson and Ng (2009), we build a single linear combination from a subset of the first eight estimated principal components, $\widehat{\mathbf{G}}_t = [\hat{g}_{1,t}, \hat{g}_{1,t}^3, \hat{g}_{3,t}, \hat{g}_{4,t}, \hat{g}_{8,t}]$ to obtain the LN factor[13]

$$LN_t^h = \hat{\boldsymbol{\lambda}}^{h\prime} \widehat{\mathbf{G}}_t, \tag{9}$$

where $\hat{\boldsymbol{\lambda}}$ is obtained from the projection

$$\frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} = \lambda_0^h + \lambda_1^h \hat{g}_{1,t} + \lambda_2^h \hat{g}_{1,t}^3 + \lambda_3^h \hat{g}_{3,t} + \lambda_4^h \hat{g}_{4,t} + \lambda_5^h \hat{g}_{8,t} + \overline{\eta}_{t+h/12}. \tag{10}$$

Panel B in Table 1 presents summary statistics for the Fama-Bliss forward spreads along with the CP and LN factors. The Fama-Bliss forward spreads are strongly positively autocorrelated with first-order autocorrelation coefficients around 0.90. The CP and LN factors are far less autocorrelated with first-order autocorrelations of 0.67 and 0.41, respectively.

Panel C shows that the Fama-Bliss spreads are strongly positively correlated. In turn, these spreads are positive correlated with the CP factor, with correlations around 0.5, but are uncorrelated with the LN factor. The LN factor captures a largely orthogonal component in relation to the other predictors. For example, its correlation with CP is only 0.18. It is also less persistent than the FB and CP factors.

3 Return Prediction Models and Estimation Methods

We next introduce the return prediction models and describe the estimation methods used in the paper.

3.1 Model specifications

Our analysis considers the three prediction variables described in the previous section. Specifically, we consider three univariate models, each of which includes one of these three factors, three bivariate models that includes two of the three predictors, and, finally, a model that includes all three predictors. This produces a total of seven different models:

1. Fama-Bliss (FB) univariate

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 fs_t^{(n,h)} + \varepsilon_{t+h/12}. \tag{11}$$

[13] Ludvigson and Ng (2009) selected this particular combination of factors using the Schwarz information criterion.

2. Cochrane-Piazzesi (CP) univariate

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 CP_t^h + \varepsilon_{t+h/12}. \tag{12}$$

3. Ludvigson-Ng (LN) univariate

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 LN_t^h + \varepsilon_{t+h/12}. \tag{13}$$

4. Fama-Bliss and Cochrane-Piazzesi factors (FB-CP)

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 fs_t^{(n,h)} + \beta_2 CP_t^h + \varepsilon_{t+h/12}. \tag{14}$$

5. Fama-Bliss and Ludvigson-Ng factors (FB-LN)

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 fs_t^{(n,h)} + \beta_2 LN_t^h + \varepsilon_{t+h/12}. \tag{15}$$

6. Cochrane-Piazzesi and Ludvigson-Ng factors (CP-LN)

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 CP_t^h + \beta_2 LN_t^h + \varepsilon_{t+h/12}. \tag{16}$$

7. Fama-Bliss, Cochrane-Piazzesi and Ludvigson-Ng predictors (FB-CP-LN)

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 fs_t^{(n,h)} + \beta_2 CP_t^h + \beta_3 LN_t^h + \varepsilon_{t+h/12}. \tag{17}$$

These models are in turn compared against the Expectation Hypothesis benchmark

$$rx_{t+h/12}^{(n)} = \beta_0 + \varepsilon_{t+h/12}, \tag{18}$$

that assumes no predictability. In each case $n \in \{2, 3, 4, 5\}$.

A large literature on stock return predictability finds evidence of a small but persistent predictable component in stock returns. Recent contributions to this literature have found that it is important to account for two features. First, return volatility varies over time and time varying volatility models fit the data far better than constant volatility models; see, e.g., Johannes et al. (2014) and Pettenuzzo et al. (2013). Stochastic volatility models can also account for fat tails—a feature that is clearly present in the monthly returns data (see Table 1). Second, the parameters of return predictability models are not stable over time but appear to undergo change; see Paye and Timmermann (2006), Dangl and Halling (2012) and Johannes et al. (2014).

To account for these features in the context of bond return predictability we consider four classes of models: (i) constant coefficient models with constant volatility; (ii) constant coefficient models with stochastic volatility; (iii) time-varying parameter models with constant volatility; and (iv) time-varying parameter models with stochastic volatility.

The constant coefficient, constant volatility model serves as a natural starting point for the out-of-sample analysis. There is no guarantee that the more complicated models with stochastic volatility and time-varying regression coefficients are capable of producing better out-of-sample forecasts since their parameters may be imprecisely estimated.

To estimate the models we adopt a Bayesian approach similar to that used in the literature on stock return predictability by studies such as Dangl and Halling (2012), Johannes et al. (2014), and Pettenuzzo et al. (2013).

Our Bayesian approach affords several advantages over the conventional estimation methods adopted by previous studies of bond return predictability. First, imprecisely estimated parameters is a big issue in the return predictability literature and so it is important to account for parameter uncertainty as is explicitly done by the Bayesian approach. Second, portfolio allocation analysis requires estimating not only the conditional mean, but also the conditional variance (under mean-variance preferences) or the full predictive density (under power utility) of returns. This is again accomplished by our method since the (posterior) predictive return distribution is the natural focus of the analysis. Third, as we shall see in Section 4, our approach also allows us to handle model uncertainty by averaging across models.

We next describe our estimation approach for each of the four classes of models. To ease the notation, for the remainder of the paper we drop the notation $t + h/12$ and replace $h/12$ with 1, with the understanding that the definition of a period will be different depending on the data frequency.

3.2 Constant Coefficients and Constant Volatility Model

The linear model projects bond excess returns $rx_{\tau+1}^{(n)}$ on a set of lagged predictors, $\mathbf{x}_{\tau}^{(n)}$:

$$\begin{aligned} rx_{\tau+1}^{(n)} &= \mu + \boldsymbol{\beta}'\mathbf{x}_{\tau}^{(n)} + \varepsilon_{\tau+1}, \quad \tau = 1, ..., t-1, \\ \varepsilon_{\tau+1} &\sim \mathcal{N}(0, \sigma_{\varepsilon}^2). \end{aligned} \tag{19}$$

Ordinary least squares (OLS) estimation of this model is straightforward and so is not further explained. However, we also consider Bayesian estimation so we briefly describe how the prior and likelihood are specified. Following standard practice, the priors for the parameters μ and $\boldsymbol{\beta}$ in (19) are assumed to be normal and independent of σ_{ε}^2

$$\begin{bmatrix} \mu \\ \boldsymbol{\beta} \end{bmatrix} \sim \mathcal{N}\left(\underline{b}, \underline{V}\right), \tag{20}$$

where

$$\underline{b} = \begin{bmatrix} \overline{rx}_t^{(n)} \\ \mathbf{0} \end{bmatrix}, \; \underline{V} = \underline{\psi}^2 \left[\left(s_{rx,t}^{(n)} \right)^2 \left(\sum_{\tau=1}^{t-1} \mathbf{x}_{\tau}^{(n)} \mathbf{x}_{\tau}^{(n)\prime} \right)^{-1} \right], \tag{21}$$

and $\overline{rx}_t^{(n)}$ and $\left(s_{rx,t}^{(n)}\right)^2$ are data-based moments:

$$\begin{aligned}\overline{rx}_t^{(n)} &= \frac{1}{t-1}\sum_{\tau=1}^{t-1} rx_{\tau+1}^{(n)},\\ \left(s_{rx,t}^{(n)}\right)^2 &= \frac{1}{t-2}\sum_{\tau=1}^{t-1}\left(rx_{\tau+1}^{(n)}-\overline{rx}_t^{(n)}\right)^2.\end{aligned}$$

Our choice of the prior mean vector $\underline{b}$ reflects the "no predictability" view that the best predictor of bond excess returns is the average of past returns. We therefore center the prior intercept on the prevailing mean of historical excess returns, while the prior slope coefficient is centered on zero.

It is common to base the priors of the hyperparameters on sample estimates, see Stock and Watson (2006) and Efron (2010). Our analysis can thus be viewed as an empirical Bayes approach rather than a more traditional Bayesian approach that fixes the prior distribution before any data are observed. We demonstrate below that, at least for a reasonable range of values, the choice of priors has little impact on our results.

In (21), $\underline{\psi}$ is a constant that controls the tightness of the prior, with $\underline{\psi} \to \infty$ corresponding to a diffuse prior on μ and $\boldsymbol{\beta}$. Our benchmark analysis sets $\underline{\psi} = n/2$. This choice means that the prior becomes looser for the longer maturities for which fundamentals-based information is likely to be more important. It also means that the posterior parameter estimates are shrunk more towards their priors for the shortest maturities which are most strongly affected by estimation error.

We assume a standard gamma prior for the error precision of the return innovation, σ_ε^{-2}:

$$\sigma_\varepsilon^{-2} \sim \mathcal{G}\left(s_{rx,t}^{-2}, \underline{v}_0\left(t-1\right)\right), \tag{22}$$

where $\underline{v}_0$ is a prior hyperparameter that controls how informative the prior is with $\underline{v}_0 \to 0$ corresponding to a diffuse prior on σ_ε^{-2}.[14] Our baseline analysis sets $\underline{v}_0 = 2/n$, again letting the priors be more diffuse the longer the bond maturity.

3.3 Stochastic Volatility Model

A large literature has found strong empirical evidence of time-varying return volatility, see Andersen et al. (2006). We accommodate such effects through a simple stochastic volatility (SV) model:

$$rx_{\tau+1}^{(n)} = \mu + \boldsymbol{\beta}'\mathbf{x}_\tau^{(n)} + \exp\left(h_{\tau+1}\right)u_{\tau+1}, \tag{23}$$

[14] Following Koop (2003), we adopt the Gamma distribution parametrization of Poirier (1995). If the continuous random variable Y has a Gamma distribution with mean $\mu > 0$ and degrees of freedom $v > 0$, we write $Y \sim \mathcal{G}(\mu, v)$ and so $E(Y) = \mu$ and $Var(Y) = 2\mu^2/v$.

where $h_{\tau+1}$ denotes the (log of) bond return volatility at time $\tau+1$ and $u_{\tau+1} \sim \mathcal{N}(0,1)$. Following common practice, the log-volatility is assumed to evolve as a driftless random walk,

$$h_{\tau+1} = h_\tau + \xi_{\tau+1}, \tag{24}$$

where $\xi_{\tau+1} \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$ and u_τ and ξ_s are mutually independent for all τ and s. While the random walk assumption for log-volatility may be unattractive from a theoretical perspective, as pointed out by Dangl and Halling (2012) this model has often been found in empirical studies to outperform models with mean-reverting volatility dynamics. The appendix explains how we estimate the SV model and set the priors.

3.4 Time varying Parameter Model

Studies such as Thornton and Valente (2012) find considerable evidence of instability in the parameters of bond return prediction models. The following time varying parameter (TVP) model allows the regression coefficients in (19) to change over time:

$$\begin{aligned} rx_{\tau+1}^{(n)} &= (\mu + \mu_\tau) + (\boldsymbol{\beta} + \boldsymbol{\beta}_\tau)' \mathbf{x}_\tau^{(n)} + \varepsilon_{\tau+1}, \quad \tau = 1, ..., t-1, \\ \varepsilon_{\tau+1} &\sim \mathcal{N}(0, \sigma_\varepsilon^2). \end{aligned} \tag{25}$$

The intercept and slope parameters $\boldsymbol{\theta}_\tau = \left(\mu_\tau, \boldsymbol{\beta}_\tau'\right)'$ are assumed to follow a random walk:[15]

$$\boldsymbol{\theta}_{\tau+1} = \boldsymbol{\theta}_\tau + \boldsymbol{\eta}_{\tau+1} \tag{26}$$

where $\boldsymbol{\theta}_1 = \mathbf{0}$, $\boldsymbol{\eta}_{\tau+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, and ε_τ and $\boldsymbol{\eta}_s$ are mutually independent for all τ and s.[16] The key parameter is $\mathbf{Q}$ which determines how rapidly the parameters $\boldsymbol{\theta}$ are allowed to change over time. We set the priors to ensure that the parameters are allowed to change only gradually. The appendix provides details on how we estimate the model and set the priors.

3.5 Time varying Parameter, Stochastic Volatility Model

Finally, we consider a general model that admits both time varying parameters and stochastic volatility (TVP-SV):

$$rx_{\tau+1}^{(n)} = (\mu + \mu_\tau) + (\boldsymbol{\beta} + \boldsymbol{\beta}_\tau)' \mathbf{x}_\tau^{(n)} + \exp(h_{\tau+1})\, u_{\tau+1}, \tag{27}$$

with

$$\boldsymbol{\theta}_{\tau+1} = \boldsymbol{\theta}_\tau + \boldsymbol{\eta}_{\tau+1}, \tag{28}$$

[15] This specification is similar to that of Dangl and Halling (2012) who find no evidence that a specification that allows for mean reversion in the parameters performs better. A more general specification with mean-reverting parameters is considered by Johannes et al. (2014).

[16] This is equivalent to writing $rx_{\tau+1}^{(n)} = \widetilde{\mu}_\tau + \widetilde{\boldsymbol{\beta}}_\tau' \mathbf{x}_\tau^{(n)} + \varepsilon_{\tau+1}$, where $\widetilde{\boldsymbol{\theta}}_1 \equiv \left(\widetilde{\mu}_1, \widetilde{\boldsymbol{\beta}}_1'\right)$ is left unrestricted.

where again $\boldsymbol{\theta}_\tau = \left(\mu_\tau, \boldsymbol{\beta}_\tau'\right)'$, and

$$h_{\tau+1} = h_\tau + \xi_{\tau+1}. \tag{29}$$

We assume that $u_{\tau+1} \sim \mathcal{N}(0,1)$, $\boldsymbol{\eta}_{\tau+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, $\xi_{\tau+1} \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$ and u_τ, $\boldsymbol{\eta}_s$ and ξ_l are mutually independent for all τ, s, and l. Again we refer to the appendix for further details on this model.

The models are estimated by Gibbs sampling methods. This allows us to generate draws of excess returns, $rx_{t+1}^{(n)}$, in a way that only conditions on a given model and the data at hand. This is convenient when computing bond return forecasts and determining the optimal bond holdings.

4 Empirical Results

This section describes our empirical results. For comparison with the existing literature, and to convey results on the importance of different features of the models such as time varying parameters and stochastic volatility, we first report results based on full-sample estimates. This is followed by an out-of-sample analysis of both the statistical and economic evidence on return predictability.

4.1 Full-sample Estimates

For comparison with extant results, Table 2 presents full-sample (1962:01-2011:12) least squares estimates for the bond return prediction models with constant parameters. While no investors could have based their historical portfolio choices on these estimates, such results are important for our understanding of how the various models work. The slope coefficients for the univariate models increase monotonically in the maturity of the bonds. With the exception of the coefficients on the CP factor in the multivariate model, they are significant across all maturities and forecasting models.[17]

Table 2 shows R^2 values around 1-2% for the model that uses FB as a predictor, 2.5% for the model that uses the CP factor and around 5% for the model based on the LN factor. These values increase to 6-8% for the multivariate models, notably smaller than those conventionally reported for the annual horizon. For comparison, at the one-year horizon we obtain R^2 values of 10-11%, 17-24%, and 14-19% for the FB, CP, and LN models, respectively. These values are in line with, if a bit weaker than, those reported in the literature. This reflects our use of an

[17] As emphasized by Cochrane and Piazzesi (2005), care has to be exercised when evaluating the statistical significance of these results due to the highly persistent FB and CP regressors. Wei and Wright (2013) find that conventional tests applied to bond excess return regressions that use yield spreads or yields as predictors are subject to considerable finite-sample distortions. However, their reverse regression approach confirms that, even after accounting for such biases, bond excess returns still appear to be predictable.

extended sample along with evidence that the regression coefficients decline towards zero at the end of the sample.

The extent of time variations in the parameters of the three-factor FB-CP-LN model is displayed in Figure 2. When interpreting the plots it should be recalled that we set the priors so the parameters are only allowed to change slowly. This ensures that the parameter estimates do not get dominated by noise. The coefficients on both the FB forward spread and the LN macrofactor in the TVP model increase systematically up to around 1985 before starting a gradual decline. Conversely, the coefficient on the CP factor is quite low during the early sample period but increases towards the end.

An advantage of our approach is its ability to deal with parameter estimation error. To get a sense of the importance of this issue, Figure 3 plots full-sample posterior densities of the regression coefficients for the three-factor model that uses the FB, CP and LN factors as predictors. The spread of the densities in this figure shows the considerable uncertainty surrounding the parameter estimates even at the end of the sample. As expected, parameter uncertainty is greatest for the TVP and TVP-SV models which allow for the greatest amount of flexibility–clearly this comes at the cost of less precisely estimated parameters. The SV model generates more precisely estimated regression coefficients than the constant volatility benchmark, reflecting the tendency of this model to reduce the weight on observations in highly volatile periods.

The effect of such parameter uncertainty on the predictive density of bond excess returns is depicted in Figure 4. This figure evaluates the univariate LN model at the mean of this predictor, plus or minus two times its standard deviation. The TVP and TVPSV models imply a greater dispersion for bond returns and their densities shift further out in the tails as the predictor variable moves away from its mean. The four models clearly imply very different probability distributions for bond returns and so can be expected to result in different implications when used by investors to form portfolios.

Figure 5 plots the time series of the posterior means and volatilities of bond excess returns for the FB-CP-LN model. Mean excess returns (top panel) vary substantially during the sample, peaking during the early eighties, nineties and again during 2008. Stochastic volatility effects (bottom panel) also appear to be empirically important. The conditional volatility is very high during 1979-1982, while subsequent spells with above-average volatility are more muted and short-lived. Interestingly, there are relatively long spells with below-average conditional volatility such as during the late nineties and mid-2000s.

4.2 Out-of-sample Analysis

To gauge the real-time value of the bond return prediction models, following Ludvigson and Ng (2009) and Thornton and Valente (2012), we next conduct an out-of-sample forecasting

experiment.[18] This experiment relies on information up to period t to compute return forecasts for period $t+1$ and uses an expanding estimation window. Notably, when constructing the CP and LN factors we also restrict our information to end at time t. Hence, we re-estimate each period the principal components and the regression coefficients in (7) and (10).

We use 1962:01-1989:12 as our initial warm-up estimation sample and 1990:01-2011:12 as the forecast evaluation period. As before, we set $n = 2, 3, 4, 5$ and so predict 2, 3, 4, and 5-year bond returns in excess of the one-month T-bill rate.

The predictive accuracy of the bond excess return forecasts is measured relative to recursively updated forecasts from the expectations hypothesis (EH) model (18) that projects excess returns on a constant. Specifically, at each point in time, we obtain draws from the predictive densities of the benchmark model and the models with time-varying predictors. For a given bond maturity n, we denote draws from the predictive density of the EH model, given the information set at time t, $\mathcal{D}^t = \{rx_{\tau+1}^{(n)}\}_{\tau=1}^{t-1}$, by $\left\{rx_{t+1}^{(n),j}\right\}$, $j = 1, ..., J$. Similarly, draws from the predictive density of any of the other models (labeled model i) given $\mathcal{D}^t = \{rx_{\tau+1}^{(n)}, \mathbf{x}_\tau^{(n)}\}_{\tau=1}^{t-1} \cup \mathbf{x}_t^{(n)}$ are denoted $\left\{rx_{t+1,i}^{(n),j}\right\}$, $j = 1, ..., J$.[19]

For the linear constant parameter, constant volatility model, return draws are obtained by applying a Gibbs sampler to

$$p\left(rx_{t+1}^{(n)}\middle|\,\mathcal{D}^t\right) = \int_{\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2}} p\left(rx_{t+1}^{(n)}\middle|\,\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\mathcal{D}^t\right) p\left(\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2}\middle|\,\mathcal{D}^t\right) d\mu d\boldsymbol{\beta} d\sigma_\varepsilon^{-2}. \tag{30}$$

Return draws for the most general TVP-SV model are obtained from the predictive density[20]

$$\begin{aligned} p\left(rx_{t+1}^{(n)}\middle|\,\mathcal{D}^t\right) \;=\; & \int_{\mu,\boldsymbol{\beta},\boldsymbol{\theta}^{t+1}\mathbf{Q},h^{t+1},\sigma_\xi^{-2}} p\left(rx_{t+1}^{(n)}\middle|\,\boldsymbol{\theta}_{t+1},h_{t+1},\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right) \\ & \times p\left(\boldsymbol{\theta}_{t+1},h_{t+1}\middle|\,\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right) \\ & \times p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},h^t,\sigma_\xi^{-2}\middle|\,\mathcal{D}^t\right) d\mu d\boldsymbol{\beta} d\boldsymbol{\theta}^{t+1} d\mathbf{Q} dh^{t+1} d\sigma_\xi^{-2}, \end{aligned} \tag{31}$$

where $h^{t+1} = (h_1, ..., h_{t+1})$ and $\boldsymbol{\theta}^{t+1} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{t+1})$ denote the sequence of conditional variance states and time varying regression parameters up to time $t+1$, respectively. Draws from the SV and TVP models are obtained as special cases of (31). All Bayesian models integrate out uncertainty about the parameters. Thornton and Valente (2012) use shrinkage methods to

[18] Out-of-sample analysis also provides a way to guard against overfitting. Duffee (2010) shows that in-sample overfitting can generate unrealistically high Sharpe ratios.

[19] We run the Gibbs sampling algorithms recursively for all time periods betweeen 1990:01 and 2011:12. At each point it time, we retain 1,000 draws from the Gibbs samplers after a burn-in period of 500 iterations. For the TVP, SV, and TVP-SV models we run the Gibbs samplers five times longer while at the same time thinning the chains by keeping only one in every five draws, thus effectively eliminating any autocorrelation left in the draws. Additional details on these algorithms are presented in the appendix.

[20] For each draw retained from the Gibbs sampler, we produce 100 draws from the corresponding predictive densities.

accommodate uncertainty in mean parameters but do not consider uncertainty about covariance parameters. Moreover, their approach is not easily generalized to settings with stochastic volatility and time varying parameters.

4.2.1 Out-of-sample Forecasts

Although our models generate a full predictive distribution for bond returns it is insightful to also report results based on conventional point forecasts. These are used extensively in the literature on stock return predictability and are reported by Ludvigson and Ng (2009) for bond returns. To obtain point forecasts we first compute the posterior mean from the densities in (30) and (31). We denote these by $\overline{rx}_{t,EH}^{(n)} = \frac{1}{J}\sum_{j=1}^{J} rx_t^{(n),j}$ and $\overline{rx}_{t,i}^{(n)} = \frac{1}{J}\sum_{j=1}^{J} rx_{t,i}^{(n),j}$, for the EH and alternative models, respectively. Using such point forecasts, we obtain the corresponding forecast errors as $e_{t,EH}^{(n)} = rx_t^{(n)} - \overline{rx}_{t,EH}^{(n)}$ and $e_{t,i}^{(n)} = rx_t^{(n)} - \overline{rx}_{t,i}^{(n)}$, $t = \underline{t}, ..., \bar{t}$, where $\underline{t} = 1990:01$ and $\bar{t} = 2011:12$ denote the beginning and end of the forecast evaluation period.

Following Campbell and Thompson (2008), we compute the out-of-sample R^2 of model i relative to the EH model as

$$R_{OoS,i}^{(n)2} = 1 - \frac{\sum_{\tau=\underline{t}}^{\bar{t}} e_{\tau,i}^{(n)2}}{\sum_{\tau=\underline{t}}^{\bar{t}} e_{\tau,EH}^{(n)2}}. \tag{32}$$

Positive values of this statistic suggest evidence of time-varying return predictability.

Table 3 reports R_{OoS}^2 values for the OLS, linear, SV, TVP and TVP-SV models across the four bond maturities. For the two-year maturity we find little evidence that models estimated by OLS are able to improve on the predictive accuracy of the EH model. Conversely, four of the seven linear models estimated using our Bayesian approach generate significantly more accurate forecasts at the 1% significance level, with another two models being significant at the 10% level, using the test for equal predictive accuracy suggested by Clark and West (2007). The SV models generate more accurate forecasts for four out of seven models at the 1% significance level with three of these coming out with an R_{OoS}^2 value above 5%. The TVP models generate similarly significant results, although the associated R_{OoS}^2 values are generally smaller than those for the SV models. The results for the TVP-SV models generally fall between those for the SV and TVP models that they nest.

While the OLS models fare considerably better for the longer bond maturities, the ability of the linear Bayesian model to generate accurate forecasts does not appear to depend as strongly on the maturity. Moreover, the Bayesian approach performs notably better than its OLS counterpart, particularly for the multivariate models.

Comparing results across predictor variables, the univariate CP model is never found to improve the predictive accuracy even among the Bayesian models and so performs the worst. Moreover, there is only modest evidence that the CP variable, when added to any of the other

predictors, results in improved performance. Conversely, the FB and LN two-factor model performs best across the four maturities.

Ranking the different specifications, we find that the SV models produce the most accurate point forecasts for the shortest maturity (2 years), while the TVP models generate the most accurate forecasts for the four and five year maturities. The results for the TVP-SV model generally fall between those obtained for the separate SV and TVP models. These results suggest that the more sophisticated models that allow for time varying parameters and time varying volatility manage to produce better out-of-sample forecasts than simple constant parameter, constant volatility models.

To identify which periods the models perform best, following Welch and Goyal (2008), we use the out-of-sample forecast errors to compute the difference in the cumulative sum of squared errors (SSE) for the EH model versus the ith model:

$$\Delta CumSSE_{t,i}^{(n)} = \sum_{\tau=\underline{t}}^{t} \left(e_{\tau,EH}^{(n)}\right)^2 - \sum_{\tau=\underline{t}}^{t} \left(e_{\tau,i}^{(n)}\right)^2. \tag{33}$$

Positive and increasing values of $\Delta CumSSE_t$ suggest that the model with time-varying return predictability generates more accurate point forecasts than the EH benchmark.

Figure 6 plots $\Delta CumSSE_t$ for the three univariate models and the three factor model, assuming a two-year bond maturity. These plots show periods during which the various models perform well relative to the EH model–periods where the lines are increasing and above zero–and periods where the models underperform against this benchmark–periods with decreasing graphs. The univariate FB model performs quite poorly due to spells of poor performance in 1994, 2000 and, again, in 2008, while the CP model underperforms between 1993 and 2005. In contrast, except for a few isolated months in 2002, 2008 and 2009, the LN model consistently beats the EH benchmark up to 2010, at which point its performance flattens against the EH model. A similar performance is seen for the multivariate model.

The predictive accuracy measures in (32) and (33) ignore information on the full probability distribution of returns. To evaluate the accuracy of the density forecasts obtained in (30) and (31), we use the log predictive score. This is commonly viewed as the broadest measure of accuracy of density forecasts, see, e.g., Geweke and Amisano (2010). At each point in time t, the log predictive score is obtained by taking the natural log of the predictive densities (30)–(31) evaluated at the observed bond excess return, $rx_t^{(n)}$, denoted by $LS_{t,EH}$ and $LS_{t,i}$ for the EH and alternative models, respectively.

Table 4 reports the average log-score differential for each of our models, again measured relative to the EH benchmark.[21] The results show that the linear model performs significantly

[21] To test if the differences in forecast accuracy are significant, we follow Clark and Ravazzolo (2014) and apply the Diebold and Mariano (1995) t-test for equality of the average log-scores based on the statistic $\overline{LS}_i =$

better than the EH benchmark across almost all variable choices for the 3-5 year bond maturities. For the 2-4 year bond maturities the evidence against the EH model is even stronger when we turn to the SV model. Interestingly, however, this model produces relatively weak rejections of the EH model for the five-year maturity. In unreported results that compare the SV models to the linear models, we find that the SV model is strongly preferred for the three shortest maturities ($n = 2, 3, 4$) but not for the longest maturity ($n = 5$).

The TVP model produces consistent, if more modest, improvements in the log-score of the EH model, but generally performs slightly worse than the linear model on this criterion due to the greater uncertainty surrounding the density forecasts for this model. Conversely, the TVPSV model performs better than the linear model on this criterion.[22]

Figure 7 supplements Table 4 by showing the cumulative log score (LS) differentials between the EH model and the ith model, computed analogously to (33) as

$$\Delta CumLS_{t,i} = \sum_{\tau=\underline{t}}^{t} \left[LS_{\tau,i} - LS_{\tau} \right]. \tag{34}$$

The dominant performance of the density forecasts generated by the SV models is clear from these plots. In contrast, the linear and TVP models offer only modest improvements over the EH benchmark by this measure.

4.3 Robustness to Choice of Priors

Choice of priors can always be debated in Bayesian analysis, so we conduct a sensitivity analysis with regard to two of the priors, namely $\underline{\psi}$ and $\underline{v}_0$, which together control how informative the baseline priors are. Our first experiment sets $\underline{\psi} = 5$ and $\underline{v}_0 = 1/5$. This choice corresponds to using more diffuse priors than in the baseline scenario. Compared with the baseline prior, this prior produces worse results (lower out-of-sample R^2 values) for the two shortest maturities ($n = 2, 3$), but stronger results for the longest maturities ($n = 4, 5$).

Our second experiment sets $\underline{\psi} = 0.5, \underline{v}_0 = 5$, corresponding to tighter priors. Under these priors, the results improve for the shorter bond maturities but get weaker at the longest maturities. In both cases, the conclusion that the best prediction models dominate the EH benchmark continues to hold even for such large shifts in priors.

$\frac{1}{\overline{t}-\underline{t}+1} \sum_{\tau=\underline{t}}^{\overline{t}} (LS_{\tau.i} - LS_{\tau,EH})$. The p-values for this statistic are based on t-statistics computed with a serial correlation-robust variance, using the pre-whitened quadratic spectral estimator of Andrews and Monahan (1992). Monte Carlo evidence in Clark and McCracken (2011) indicates that, with nested models, the Diebold-Mariano test compared against normal critical values can be viewed as a somewhat conservative test for equal predictive accuracy in finite samples. Since all models considered here nest the EH benchmark, we report p-values based on one-sided tests, taking the nested EH benchmark as the null and the nesting model as the alternative.

[22] Comparing the predictive likelihood of the SV model to that of the linear specification, we find (in unreported results) that the SV model produces significantly better results for the 2-4 year bond maturities.

5 Economic Value and Drivers of Bond Return Predictability

We next discuss the economic value and drivers of the evidence on bond return predictability established in the previous section. We first consider the economic value of our out-of-sample bond return forecasts to an investor with power utility. Next, we analyze the link between the economic cycle and bond return predictability. Finally, we explore how our bond return forecasts are correlated with drivers of time varying bond risk premia.

5.1 Economic Value of Return Forecasts

So far our analysis concentrated on statistical measures of predictive accuracy. It is important to evaluate the extent to which the apparent gains in predictive accuracy translate into better investment performance. In fact, for an investor with mean-variance preferences, Thornton and Valente (2012) find that improvements in the statistical accuracy of bond return forecasts do not imply improved portfolio performance.

We consider the asset allocation decisions of an investor that selects the weight, $\omega_t^{(n)}$, on a risky bond with n periods to maturity versus a one-month T-bill that pays the riskfree rate, $\tilde{y}_t = y_t^{(1/12)}$. The investor has power utility and coefficient of relative risk aversion A:

$$U\left(\omega_t^{(n)}, rx_{t+1}^{(n)}\right) = \frac{\left[\left(1-\omega_t^{(n)}\right)\exp\left(\tilde{y}_t\right) + \omega_t^{(n)}\exp\left(\tilde{y}_t + rx_{t+1}^{(n)}\right)\right]^{1-A}}{1-A}, \quad A > 0. \tag{35}$$

Using all information at time t, $\mathcal{D}^t$, to evaluate the predictive density of $rx_{t+1}^{(n)}$, the investor solves the optimal asset allocation problem

$$\omega_t^{(n)*} = \arg\max_{\omega_t^{(n)}} \int U\left(\omega_t^{(n)}, rx_{t+1}^{(n)}\right) p\left(rx_{t+1}^{(n)} \middle| \mathcal{D}^t\right) drx_{t+1}^{(n)}. \tag{36}$$

The integral in (36) can be approximated by generating a large number of draws, $rx_{t+1,i}^{(n),j}$, $j = 1, .., J$, from the predictive densities specified in (30) and (31). For each of the candidate models, i, we approximate the solution to (36) by

$$\widehat{\omega}_{t,i}^{(n)} = \arg\max_{\omega_{t,i}^{(n)}} \frac{1}{J}\sum_{j=1}^{J}\left\{\frac{\left[\left(1-\omega_{t,i}^{(n)}\right)\exp\left(\tilde{y}_t\right) + \omega_{t,i}^{(n)}\exp\left(\tilde{y}_t + rx_{t+1,i}^{(n),j}\right)\right]^{1-A}}{1-A}\right\}. \tag{37}$$

The resulting sequences of portfolio weights $\left\{\widehat{\omega}_{t,EH}^{(n)}\right\}$ and $\left\{\widehat{\omega}_{t,i}^{(n)}\right\}$ are next used to compute realized utilities. For each model, i, we convert these into certainty equivalent returns (CER), i.e., values that equate the average utility of the EH model with the average utility of any of the alternative models.

We consider two different sets of assumptions about the portfolio weights. The first scenario restricts the weights on the risky bonds to the interval $[0, 0.99]$ to ensure that the expected

utility is finite even with an unbounded return distribution. See Geweke (2001) and Kandel and Stambaugh (1996) for a discussion of this point. The second scenario leaves the portfolio weights unrestricted and instead restricts the bond returns to fall between -100% and 100%. This follows the argument in Johannes et al. (2014) that such a procedure prevents the expected utility from becoming unbounded.

In both cases we set the coefficient of relative risk aversion to $A = 10$, a value higher than normally considered. Our choice reflects the high Sharpe ratios observed for the bond portfolios during our sample—see Table 1. For lower values of A, this causes the weights on the risky bonds to almost always hit the upper bound (0.99) of the first scenario for both the EH and time varying predictability models and so does not allow us to differentiate between these models. However, provided that we do not impose too tight limits on the bond portfolio weights (Scenario 2), informative results can still be obtained for lower values of A (e.g., $A = 5$), as we discuss below.

5.1.1 Empirical Results

Table 5 shows annualized CER values computed relative to the EH model so positive values indicate that the time varying predictability models perform better than the EH model.

For the scenario with the weights constrained to $[0, 0.99]$ (Panels A-D), the CER values generally increase with the bond maturity. The highest CER values are generally found for the two-factor FB-LN and three-factor FB-CP-LN models. For these models the CER values increase from around 0.1% ($n = 2$) to 0.5% ($n = 3$) and 1%-1.3% for the longest bond maturity. Ranked across forecasting models the best results are achieved by the SV or TVP-SV models. The SV model is best for roughly half the models for maturities $n = 2, 3$ years, while the TVP-SV model accomplishes a similar level of performance for $n = 4, 5$. To test if the annualized CER-values are statistically greater than zero we use a one-sided Diebold-Mariano test.[23] Except for the two-year bond, the CER values of most models are significantly higher than those generated by the EH benchmark, provided that the LN predictor is included.

Figure 8 plots cumulative CER values, computed relative to the EH benchmark, for the FB-CP-LN three-factor model and assuming $\hat{\omega}_t^{(n)} \in [0, 0.99]$. These graphs parallel the cumulated sum of squared error difference plots in (33), the key difference being that they show the cumulated risk-adjusted gains from using a particular model instead of the EH model. For the two-year bond maturity we uncover little evidence that the models improve on the economic

[23] Specifically, we estimate the regression $u_{i,t+1}^{(n)} - u_{EH,t+1}^{(n)} = \alpha^{(n)} + \epsilon_{t+1}$ where

$$u_{i,t+1}^{(n)} = \frac{1}{1-A}\left[\left(1-\omega_{t,i}^{(n)}\right)\exp\left(\tilde{y}_t\right) + \omega_{t,i}^{(n)}\exp\left(\tilde{y}_t + rx_{t+1}^{(n)}\right)\right]^{1-A},$$

and

$$u_{EH,t+1}^{(n)} = \frac{1}{1-A}\left[\left(1-\omega_{t,EH}^{(n)}\right)\exp\left(\tilde{y}_t\right) + \omega_{t,EH}^{(n)}\exp\left(\tilde{y}_t + rx_{t+1}^{(n)}\right)\right]^{1-A},$$

and test if $\alpha^{(n)}$ equals zero.

performance of the EH model. Conversely, as the bond maturity increases from three through five years, we note improved economic performance. Interestingly, the models' best performance seems to be concentrated from 1990 to 1993 and from 2001 onwards. For the five-year bond, the cumulated CER value at the end of the sample exceeds 30 percent for all models.

Turning to the case with unconstrained portfolio weights (Panels E-H in Table 5), the CER values generally increase substantially, notably for the two-year bond maturity. For example, for the SV model with three predictors the CER increases from 0.14% to 5.42% (two-year bond) and from 1.21% to 1.92% (five-year maturity), showing that the SV model is particularly affected by the tight constraints on the portfolio weights under the first scenario. Similar improvements are observed for the TVP model, whereas the evidence is more mixed for the TVP-SV model which is most strongly affected by parameter uncertainty and so sometimes benefits from constraints on the portfolio weights (see Jagannathan and Ma (2003)). Interestingly, in the absence of tight constraints on the portfolio weights, we also obtain significant improvements in the CER values of the SV and TVP-SV Bayesian models for smaller values of the coefficient of relative risk aversion, e.g., $A = 5$ (results not shown here).

Comparing results across the different specifications, the SV model generally performs best. Its improvements relative to the linear model are particularly large in the case with unconstrained weights for the two year bond maturity where we find CER gains of 2-4% per year. However, the CER gains continue to be significantly higher than those obtained from the linear model for the three and four year bonds. For the multivariate models the TVP model also produces higher CER values than the linear model.

We conclude from these results that there is strong statistical and economic evidence that the returns on 2-5 year bonds can be predicted using predictor variables proposed in the literature. Moreover, the best performing models do not assume constant parameters but allow for a time varying mean and volatility dynamics.

Our results are very different from those reported by Thornton and Valente (2012). These authors find that statistical evidence of out-of-sample return predictability fails to translate into an ability for investors to use return forecasts in a way that generates higher out-of-sample average utility than forecasts from the EH model. Notably, whereas we find that accounting for time varying parameters and stochastic volatility in many cases improves bond portfolio performance, Thornton and Valente (2012) find that the Sharpe ratios of their bond portfolios decrease when accounting for such effects through rolling window estimation.

Besides differences in modeling approaches, a reason for such differences is the focus of Thornton and Valente (2012) on 12-month bond returns, whereas we use monthly bond returns. To address the importance of the return horizon, we repeat the out-of-sample analysis using non-overlapping quarterly and annual returns data. Compared with the monthly results, the quarterly and annual R^2 values decline somewhat. At the quarterly horizon the univariate FB

and LN models, along with the bivariate FB-LN model, continue to perform well across the four bond maturities. The LN and FB-LN models also perform well at the annual horizon, particularly for the bonds with shorter maturities ($n = 2, 3$). The associated CER values continue to be positive and, in most cases, significant at the quarterly horizon, but are substantially smaller at the annual horizon. These findings indicate a fast moving predictable component in bond returns that is missed when using longer return horizons and so help explain the difference between our results and those of Thornton and Valente (2012) and Dewachter et al. (2014).

The setup of Sarno et al. (2014) is closest to that adopted here as they also consider results for one-month returns and still obtain negative economic values from using their time-varying bond return forecasts compared with the EH model. Such differences in results reflect (i) different modeling assumptions. Sarno et al. (2014) compute expected excess returns in the context of an affine term structure model and also do not consider stochastic volatility or time-varying parameters, (ii) different predictor variables–Sarno et al. (2014) use latent state variables extracted from their term structure model to predict bond excess returns, and, (iii) different estimation methodologies–Sarno et al. (2014) do not follow the same Bayesian methodology that we use here and thus ignore parameter uncertainty.

5.2 Cyclical variations in bond return predictability

Recent studies such as Rapach et al. (2010), Henkel et al. (2011) and Dangl and Halling (2012) report that predictability of stock returns is concentrated in economic recessions and is largely absent during expansions. This finding is important since it suggests that return predictability is linked to cyclical variations and that time varying risk premia may be important drivers of expected returns.

To see if bond return predictability varies over the economic cycle, we split the data into recession and expansion periods using the NBER recession indicator with recessions labeled '1' while expansions are labeled '0'. Table 6 uses the full-sample parameter estimates shown in Figure 2, but computes R^2 values separately for the recession and expansion samples. We use full-sample information because there are only three recessions in our out-of-sample period, 1990-2011.

Table 6 shows that the R^2 values are generally much higher during recessions than in expansions. This finding is consistent with the findings for stock market returns as indicated by the earlier references. Moreover, it is robust across model specifications and predictor variables, the one exception being the univariate FB model for which return predictability actually is stronger during expansions. Conversely, note that the R^2 values are particularly high in recessions for the TVP models that include the LN variable.

To test if the differences in R^2 values are statistically significant, we conduct a simple bootstrap test that exploits the monotonic relationship between the mean squared prediction error

(MSE) of the forecasting model, measured relative to that of the EH model, and the R^2 measure in (32). Specifically, we test the null that the predictive accuracy of a given prediction model (measured relative to the EH benchmark) is the same across recessions and expansions, against the one-sided alternative that the relative MSE is higher in expansions,

$$\begin{aligned} H_0: &\quad E[\underbrace{e^2_{EH,0} - e^2_{i,0}}_{\Delta_0}] = E[\underbrace{e^2_{EH,1} - e^2_{i,1}}_{\Delta_1}] \\ H_1: &\quad E[e^2_{EH,0} - e^2_{i,0}] < E[e^2_{EH,1} - e^2_{i,1}]. \end{aligned} \tag{38}$$

Here e_{EH} and e_i are the forecast errors under the EH and model i, respectively, and the subscript refers to expansions (0) and recessions (1). By computing a particular model's MSE relative to the MSE of the EH model in the same state we control for differences in bond return variances in recessions versus expansions. Our test uses a bootstrap based on the frequency with which $\Delta_0 - \Delta_1$ is smaller than 10,000 counterparts bootstrapped under the null of $\Delta_0 = \Delta_1$.[24]

Outcomes from this test are indicated by stars in the recession columns of Table 6. We find that not only is the model fit of most bond return prediction models generally better in recessions than in expansions, but this difference is highly statistically significant in most cases.

Large differences between bond return predictability in recessions and expansions are also observed in the out-of-sample period 1990-2011. However, in this case we do not have a large enough number of recessions for the test of equal predictive power to have sufficient power to reject the null hypothesis in (38).[25]

5.3 Time varying risk premia

Asset pricing models such as Campbell and Cochrane (1999) suggest that the Sharpe ratio on risky assets should be higher during recessions due to higher consumption volatility and a lower surplus consumption ratio. To see if this implication is consistent with our models, Table 7 reports Sharpe ratios for the bond portfolios computed separately for recession and expansion periods. Following authors such as Henkel et al. (2011) these results are again based on the full sample to ensure enough observations in recessions. For most models the Sharpe ratios are substantially higher during recessions than in expansions.

[24] The p-value for the test is computed as follows: i) impose the null of equal-predictability across states i.e., compute $\hat{\Delta}_0 = \Delta_0 - \hat{\mu}(\Delta_0)$ and $\hat{\Delta}_1 = \Delta_1 - \hat{\mu}(\Delta_1)$; ii) estimate the distribution under the null by using an i.i.d. bootstrap, generate B bootstrap samples from $\hat{\Delta}_0$ and $\hat{\Delta}_1$ and for each of these compute $J^b = \mu(\hat{\Delta}_0^b) - \mu(\hat{\Delta}_1^b)$; iii) compute p-values as $p_{val} = \frac{1}{B}\sum_{b=1}^{B} 1[J > J^b]$ where $J = \mu(\Delta_0) - \mu(\Delta_1)$ is based on the data.

[25] Engsted et al. (2013) find that bond return predictability is stronger during expansions than during recessions, concluding that return predictability displays opposite patterns in the bond and stock markets. However, they use returns on a 20-year Treasury bond obtained from Ibbotson International. As we have seen, bond return predictability strongly depends on the bond maturity and so this is likely to explain the difference between their results and ours.

To further analyze risk-based explanations of bond return predictability, we explore two important sources of bond risk premia. First, many asset pricing models suggest that investors' risk aversion should vary countercyclically, being higher around recessions and lower in expansions. To the extent that our forecasts of bond excess returns reflect time varying risk premia, we should therefore expect a negative correlation between economic growth and bond return forecasts. Second, inflation risk is likely to be an important determinant of bond return dynamics, see, e.g., Wright (2011) and Abrahams et al. (2013) and we should expect to find a positive correlation between inflation uncertainty and bond return forecasts.

Table 8 tests this implication. The table reports the contemporaneous correlations between forecasts of two-year bond excess returns and current real GDP growth (Panel A), inflation (Panel B), real GDP growth uncertainty (Panel C) and inflation uncertainty (Panel D). Real GDP growth is computed as $\Delta \log(GDP_{t+1})$, where GDP_{t+1} is the real gross domestic product, while inflation is computed as $\Delta \log(CPI_{t+1})$, where CPI is the consumer price index for all urban consumers. We measure real GDP growth and inflation uncertainty using the cross-sectional dispersion (the difference between the 75th percentile and the 25th percentile) in real GDP and CPI one quarter ahead forecasts, respectively, as reported by the Survey of Professional Forecasters maintained by the Philadelphia Federal Reserve. An advantage of this measure is that it affords a model-free approach.

The correlations between out-of-sample bond excess return forecasts, on the one hand, and current GDP growth or inflation, on the other, are negative and, in most cases, highly significant. Thus lower economic growth and reduced inflation appear to be associated with expectations of higher bond excess returns.

Turning to the uncertainty measures, we find a strongly positive and, in most cases, highly significant correlation between uncertainty about economic growth and future inflation, on the one hand, and expected bond excess returns on the other. This is consistent with our bond excess return forecasts being driven, at least in part, by time-varying inflation risk premia. Correlations are particularly strong for the models that include the LN macro factor which can be expected to be particularly sensitive to the economic cycle.

6 Model Combinations

In addition to parameter uncertainty, investors face model uncertainty. This raises the question whether, in real time, investors could have selected forecasting models that would have generated accurate forecasts. Model uncertainty would not be a concern if all prediction models produced improvements over the EH benchmark. However, as we have seen in the empirical analysis, there is a great deal of heterogeneity across the models' predictive performance. To address this issue, we turn to model combination. Model combinations form portfolios of individual prediction

models with weights reflecting the models' historical performance. The better a model's fit relative to its complexity, the larger its weight. Similar to diversification benefits obtained for asset portfolios, model combination tends to stabilize forecasts relative to forecasts generated by individual return prediction models.

A second reason for our interest in model combinations is that studies on predictability of stock returns such as Rapach et al. (2010), Dangl and Halling (2012), and Pettenuzzo et al. (2013) find that combinations improve on the average performance of the individual models. This result has only been established for stock returns, however. To see if it carries over to bond returns, we consider three different combination schemes applied to the seven different choices of predictors (11)-(17) and the linear, SV, and TVP models introduced in Section 3, for a total of $N = 21$ possible models.[26]

6.1 Combination Schemes

We begin by considering the equal-weighted pool (EWP) which weighs each of the N models, M_i, equally

$$p\left(rx_{t+1}^{(n)}\middle|\,\mathcal{D}^t\right) = \frac{1}{N}\sum_{i=1}^{N} p\left(rx_{t+1}^{(n)}\middle|\,M_i,\mathcal{D}^t\right), \tag{39}$$

where $\left\{p\left(rx_{t+1}^{(n)}\middle|\,M_i,\mathcal{D}^t\right)\right\}_{i=1}^{N}$ denotes the predictive densities specified in (30) and (31).

We also consider Bayesian model averaging (BMA) weights:

$$p\left(rx_{t+1}^{(n)}\middle|\,\mathcal{D}^t\right) = \sum_{i=1}^{N} \Pr\left(M_i|\,\mathcal{D}^t\right) p\left(rx_{t+1}^{(n)}\middle|\,M_i,\mathcal{D}^t\right). \tag{40}$$

Here $\Pr\left(M_i|\,\mathcal{D}^t\right)$ denotes the posterior probability of model i, relative to all models under consideration, computed using information available at time t, $\mathcal{D}^t$. This is given by

$$\Pr\left(M_i|\,\mathcal{D}^t\right) = \frac{\Pr\left(\mathcal{D}^t\middle|\,M_i\right)\Pr\left(M_i\right)}{\sum_{j=1}^{N}\Pr\left(\mathcal{D}^t|\,M_j\right)\Pr\left(M_j\right)}. \tag{41}$$

$\Pr\left(\mathcal{D}^t\middle|\,M_i\right)$ and $\Pr\left(M_i\right)$ denote the marginal likelihood and prior probability for model i, respectively. We assume that all models are equally likely a priori and so set $\Pr\left(M_i\right) = 1/N$.[27]

A limitation of the BMA approach is that it assumes that the true prediction model is contained in the set of models under consideration. One approach that does not require this

[26] We omit the TVP-SV model from the combination analysis since it nests all the other models. Model combination is naturally viewed as an alternative to trying to come up with a single large nesting model. However, we also performed the combination analysis with the TVP-SV models included and obtained similar results to those reported here.

[27] We follow Geweke and Amisano (2010) and compute the marginal likelihoods by cumulating the predictive log scores of each model over time after conditioning on the initial warm-up estimation sample $\Pr\left(\{rx_{\tau+1}^{(n)}\}_{\tau=1}^{t-1}\middle|\,M_i\right) = \exp\left(\sum_{\tau=\underline{t}}^{t} LS_{\tau,i}\right)$.

assumption is the optimal predictive pool (OW) proposed by Geweke and Amisano (2011). This approach again computes a weighted average of the predictive densities:

$$p\left(rx_{t+1}^{(n)}\middle|\,\mathcal{D}^t\right)=\sum_{i=1}^{N}w_{t,i}^{*}\times p\left(rx_{t+1}^{(n)}\middle|\,M_i,\mathcal{D}^t\right). \tag{42}$$

The $(N\times 1)$ vector of model weights $\mathbf{w}_t^{*}=\left[w_{t,1}^{*},...,w_{t,N}^{*}\right]$ is determined by recursively solving the following maximization problem

$$\mathbf{w}_t^{*}=\arg\max_{\mathbf{w}}\sum_{\tau=1}^{t-1}\log\left[\sum_{i=1}^{N}w_i\times S_{\tau+1,i}\right], \tag{43}$$

where $S_{\tau+1,i}=\exp\left(LS_{\tau+1,i}\right)$ is the recursively computed log-score for model i at time $\tau+1$, and $\mathbf{w}_t^{*}\in[0,1]^N$. As $t\to\infty$ the weights in (43) minimize the Kullback-Leibler distance between the combined predictive density and the data generating process, see Hall and Mitchell (2007).

By recursively updating the combination weights in (40) and (43), these combination methods accommodate structural breaks or trends in the underlying dynamics.[28] This is empirically important as we shall see.

6.2 Empirical Findings

Figure 9 shows the evolution over time in the optimal combination weights for the prediction pool in (43). Regardless of the bond maturity, the linear model gets assigned little or no weight. For the two-year bond maturity the SV model assumes close to 90% of the weighting, with the remaining 10% going to the TVP model. For maturities from three through five years, the weight on the TVP model starts at 100% and decreases to about zero ($n=3$), 30% ($n=4$) and 50% ($n=5$). These plots show substantial variation over time in the amount of support offered to the individual models by the data.

Figure 10 presents the posterior probability weights on the three predictor variables computed as $p_{jt}=\sum_{i=1}^{n}w_{t,i}^{*}\mathbb{I}\{x_{ijt}=1\}$, where $w_{t,i}^{*}$ is the probability weight on model i at time t and $\mathbb{I}\{x_{ijt}=1\}$ is an indicator function that equals one for variable $j\in\{FB,CP,LN\}$ if this variable is used by model i. The plot shows that the most heavily weighted models all include the FB variable up to around 1998, after which point this variable gets a reduced weight of 60-80 percent for most of the bond maturities. The LN variable gets nearly full weight throughout most of the sample for all bond maturities. In contrast, the CP variable receives close to zero weight in the model combination.

Table 9 presents statistical and economic measures of out-of-sample forecasting performance for the three combination schemes. The optimal prediction pool generates R_{OoS}^2 values above 5%

[28] Although the TVP models account for gradually changing parameters, they do not account for more sudden shifts in the model parameters.

regardless of the bond maturity while the R^2_{OoS} values are closer to 4% for the EW combination scheme (Panel A). The BMA combination performs better than the EW combination for $n = 2$, but slightly worse for longer maturities. In all cases, the forecast combinations perform better–considerably so in the case of the optimal pool–than what one would expect from simply selecting a model at random.[29]

To see the evolution over time in the performance of the forecast combinations, the left windows in Figure 11 show the cumulative sum of squared errors for the equal-weighted pool (39), BMA weights (40) and the optimal prediction pool (43) for two and five year bond maturities. All three combination schemes outperform the EH benchmark throughout most of the sample. Moreover, the optimal prediction pool dominates the equal-weighted combination throughout almost the entire sample, including at the end.

The predictive likelihood tests shown in Panel B of Table 9 strongly reject the null of equal predictive accuracy relative to the EH model. Finally, the CER-values with constrained weights (Panel C) are quite similar for the three combination schemes, rising from about 0.1% for $n = 2$ to 1%–1.3% for $n = 5$. These CER values are similar to those achieved by the best of the individual models reported in Table 5 and, for bond maturities of three years or longer, are significantly higher than those obtained by an EH investor suggesting that model combination can be used to effectively deal with model uncertainty. The CER values for the case with unconstrained weights (Panel D) are very high, exceeding 6%/year for $n = 2$ and falling between 1.5% and 3.7% for the longer bond maturities.

The right windows in Figure 11 show cumulative CER plots for the two combination schemes, again benchmarked against the EH model. The model combinations clearly dominate the EH benchmark. Moreover, the optimal prediction pool continues to produce the best results.

7 Conclusion

We analyze predictability of excess returns on Treasury bonds with maturities ranging from two through five years. As predictors we use the forward spread variable of Fama and Bliss (1987), the Cochrane and Piazzesi (2005) combination of forward rates, and the Ludvigson and Ng (2009) macro factors. Our analysis allows for time-varying regression parameters and stochastic volatility dynamics and accounts for both parameter and model uncertainty. Using a flexible setup turns out to be important as we find significant statistical and economic gains over the constant coefficient, constant volatility models generally adopted in the existing literature.

Our findings suggest that there is evidence of both statistically and economically significant predictability in bond returns. This contrasts with the findings of Thornton and Valente (2012)

[29] The simple average of the individual models' R^2_{OoS} values are 2.51%, 2.99%, 3.09%, and 3.05% for bond maturities rising from $n = 2$ through $n = 5$ years.

who conclude that the statistical evidence on bond return predictability fails to translate into economic return predictability. Moreover, we link the evidence on return predictability to the economic cycle, finding that the degree of return predictability is significantly higher during recessions, consistent with findings reported for stock returns. Moreover, our bond return forecasts are strongly positively correlated with inflation uncertainty and negatively correlated with economic growth, consistent with time varying risk premia being an important driver of the results.

A Appendix

This appendix explains how we obtain parameter estimates for the models described in Section 3 and shows how we use these to generate predictive densities for bond excess returns. We begin by discussing the linear regression model in (19), then turn to the SV model in (23)-(24), the TVP model in (25)-(26), and the general TVPSV model in (27)-(29).

A.1 Constant coefficient, constant volatility model

The goal for the simple linear regression model is to obtain draws from the joint posterior distribution $p\left(\mu, \boldsymbol{\beta}, \sigma_{\varepsilon}^{-2} \middle| \mathcal{D}^t\right)$, where D^t denotes all information available up to time t. Combining the priors in (20)-(22) with the likelihood function yields the following posteriors:

$$\left[\begin{array}{c} \mu \\ \boldsymbol{\beta} \end{array}\right] \Bigg| \sigma_{\varepsilon}^{-2}, \mathcal{D}^t \sim \mathcal{N}\left(\overline{b}, \overline{V}\right), \tag{A-1}$$

and

$$\sigma_{\varepsilon}^{-2} \big| \mu, \boldsymbol{\beta}, \mathcal{D}^t \sim \mathcal{G}\left(\overline{s}^{-2}, \overline{v}\right), \tag{A-2}$$

where

$$\begin{aligned} \overline{V} &= \left[\underline{V}^{-1} + \sigma_{\varepsilon}^{-2} \sum_{\tau=1}^{t-1} \mathbf{x}_{\tau}^{(n)} \mathbf{x}_{\tau}^{(n)\prime}\right]^{-1}, \\ \overline{b} &= \overline{V}\left[\underline{V}^{-1}\underline{b} + \sigma_{\varepsilon}^{-2} \sum_{\tau=1}^{t-1} \mathbf{x}_{\tau}^{(n)} rx_{\tau+1}^{(n)}\right], \\ \overline{v} &= \underline{v}_0 + (t-1). \end{aligned} \tag{A-3}$$

and

$$\overline{s}^2 = \frac{\sum_{\tau=1}^{t-1}\left(rx_{\tau+1}^{(n)} - \mu - \boldsymbol{\beta}'\mathbf{x}_{\tau}^{(n)}\right)^2 + \left(s_{rx,t}^2 \times \underline{v}_0 (t-1)\right)}{\overline{v}}. \tag{A-4}$$

Gibbs sampling can be used to iterate back and forth between (A-1) and (A-2), yielding a series of draws for the parameter vector $\left(\mu, \boldsymbol{\beta}, \sigma_{\varepsilon}^{-2}\right)$. Draws from the predictive density $p\left(rx_{t+1}^{(n)} \middle| \mathcal{D}^t\right)$

can then be obtained by noting that

$$p\left(rx_{t+1}^{(n)}\middle|\,\mathcal{D}^t\right)=\int_{\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2}}p\left(rx_{t+1}^{(n)}\middle|\,\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)p\left(\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2}\middle|\,\mathcal{D}^t\right)d\mu d\boldsymbol{\beta}d\sigma_\varepsilon^{-2}. \tag{A-5}$$

A.2 Stochastic Volatility model[30]

The SV model requires specifying a joint prior for the sequence of log return volatilities, h^t, and the error precision, σ_ξ^{-2}. Writing $p\left(h^t,\sigma_\xi^{-2}\right)=p\left(h^t\middle|\,\sigma_\xi^{-2}\right)p\left(\sigma_\xi^{-2}\right)$, it follows from (24) that

$$p\left(h^t\middle|\,\sigma_\xi^{-2}\right)=\prod_{\tau=1}^{t-1}p\left(h_{\tau+1}\middle|\,h_\tau,\sigma_\xi^{-2}\right)p\left(h_1\right), \tag{A-6}$$

with $h_{\tau+1}|\,h_\tau,\sigma_\xi^{-2}\sim\mathcal{N}\left(h_\tau,\sigma_\xi^2\right)$. Thus, to complete the prior elicitation for $p\left(h^t,\sigma_\xi^{-2}\right)$, we only need to specify priors for h_1, the initial log volatility, and σ_ξ^{-2}. We choose these from the normal-gamma family as follows:

$$h_1\sim\mathcal{N}\left(\ln\left(s_{rx,t}\right),\underline{k}_h\right), \tag{A-7}$$

$$\sigma_\xi^{-2}\sim\mathcal{G}\left(1/\underline{k}_\xi,\underline{\upsilon}_\xi\right). \tag{A-8}$$

We set $\underline{k}_\xi=0.01$ and set the remaining hyperparameters in (A-7) and (A-8) at $\underline{k}_h=10$ and $\underline{\upsilon}_\xi=1$ to imply uninformative priors, thus allowing the data to determine the degree of time variation in the return volatility.

To obtain draws from the joint posterior distribution $p\left(\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2}\middle|\,\mathcal{D}^t\right)$ under the SV model, we use the Gibbs sampler to draw recursively from the following three conditional distributions:

1. $p\left(h^t\middle|\,\mu,\boldsymbol{\beta},\sigma_\xi^{-2},\mathcal{D}^t\right)$.
2. $p\left(\mu,\boldsymbol{\beta}|\,h^t,\sigma_\xi^{-2},\mathcal{D}^t\right)$.
3. $p\left(\sigma_\xi^{-2}\middle|\,\mu,\boldsymbol{\beta},h^t,\mathcal{D}^t\right)$.

We simulate from each of these blocks as follows. Starting with $p\left(h^t\middle|\,\mu,\boldsymbol{\beta},\sigma_\xi^{-2},\mathcal{D}^t\right)$, we employ the algorithm of Kim et al. (1998). Define $rx_{\tau+1}^{(n)*}=rx_{\tau+1}^{(n)}-\mu-\beta'\mathbf{x}_\tau^{(n)}$ and note that $r_{\tau+1}^{(n)*}$ is observable conditional on μ,β. Next, rewrite (23) as

$$rx_{\tau+1}^{(n)*}=\exp\left(h_{\tau+1}\right)u_{\tau+1}. \tag{A-9}$$

[30] See Pettenuzzo et al. (2013) for a description of a similar algorithm where the priors are modified to impose economic constraints on the model parameters.

Squaring and taking logs on both sides of (A-9) yields a new state space system that replaces (23)-(24) with

$$rx_{\tau+1}^{(n)**} = 2h_{\tau+1} + u_{\tau+1}^{**}, \tag{A-10}$$

$$h_{\tau+1} = h_{\tau} + \xi_{\tau+1}, \tag{A-11}$$

where $rx_{\tau+1}^{(n)**} = \ln\left[\left(rx_{\tau+1}^{(n)*}\right)^2\right]$, and $u_{\tau+1}^{**} = \ln\left(u_{\tau+1}^2\right)$, with u_{τ}^{**} independent of ξ_s for all τ and s. Since $u_{t+1}^{**} \sim \ln\left(\chi_1^2\right)$, we cannot resort to standard Kalman recursions and simulation algorithms such as those in Carter and Kohn (1994) or Durbin and Koopman (2002). To get around this problem, Kim et al. (1998) employ a data augmentation approach and introduce a new state variable $s_{\tau+1}$, $\tau = 1, .., t-1$, turning their focus on drawing from $p\left(h^t \middle| \mu, \boldsymbol{\beta}, \sigma_{\xi}^{-2}, s^t, \mathcal{D}^t\right)$ instead of $p\left(h^t \middle| \mu, \boldsymbol{\beta}, \sigma_{\xi}^{-2}, \mathcal{D}^t\right)$, where $s^t = \{s_2, s_3, ..., s_t\}$ denotes the history up to time t of the new state variable s.

The introduction of the state variable $s_{\tau+1}$ allows us to rewrite the linear non-Gaussian state space representation in (A-10)-(A-11) as a linear Gaussian state space model, making use of the following approximation,

$$u_{\tau+1}^{**} \approx \sum_{j=1}^{7} q_j \mathcal{N}\left(m_j - 1.2704, v_j^2\right), \tag{A-12}$$

where m_j, v_j^2, and q_j, $j = 1, 2, ..., 7$, are constants specified in Kim et al. (1998) and thus need not be estimated. In turn, (A-12) implies

$$u_{\tau+1}^{**} \middle| \, s_{\tau+1} = j \sim \mathcal{N}\left(m_j - 1.2704, v_j^2\right), \tag{A-13}$$

where each state has probability

$$\Pr\left(s_{\tau+1} = j\right) = q_j. \tag{A-14}$$

Conditional on s^t, we can rewrite the nonlinear state space system as follows:

$$\begin{aligned} rx_{\tau+1}^{(n)**} &= 2h_{\tau+1} + e_{\tau+1}, \\ h_{\tau+1} &= h_{\tau} + \xi_{\tau+1}, \end{aligned} \tag{A-15}$$

where $e_{\tau+1} \sim \mathcal{N}\left(m_j - 1.2704, v_j^2\right)$ with probability q_j. For this linear Gaussian state space system, we can use the algorithm of Carter and Kohn (1994) to draw the whole sequence of stochastic volatilities, h^t.

Conditional on the sequence h^t, draws for the sequence of states s^t can be obtained from

$$\Pr\left(s_{\tau+1} = j \middle| \, rx_{\tau+1}^{(n)**}, h_{\tau+1}\right) = \frac{f_h\left(rx_{\tau+1}^{(n)**} \middle| \, 2h_{\tau+1} - m_j + 1.2704, v_j^2\right)}{\sum_{l=1}^{7} f_h\left(rx_{\tau+1}^{(n)**} \middle| \, 2h_{\tau+1} - m_l + 1.2704, v_l^2\right)}. \tag{A-16}$$

Moving on to $p\left(\mu,\boldsymbol{\beta}|\, h^t,\sigma_\xi^{-2},\mathcal{D}^t\right)$, conditional on h^t it is straightforward to draw μ and $\boldsymbol{\beta}$ and apply standard results. Specifically,

$$\left[\begin{array}{c}\mu\\ \boldsymbol{\beta}\end{array}\right]\Bigg| h^t,\sigma_\xi^{-2},\mathcal{D}^t\sim\mathcal{N}\left(\overline{b},\overline{V}\right), \tag{A-17}$$

with

$$\begin{aligned}\overline{V} &= \left\{\underline{V}^{-1}+\sum_{\tau=1}^{t-1}\frac{1}{\exp\left(h_{\tau+1}\right)^2}\mathbf{x}_\tau^{(n)}\mathbf{x}_\tau^{(n)\prime}\right\}^{-1},\\ \overline{b} &= \overline{V}\left\{\underline{V}^{-1}\underline{b}+\sum_{\tau=1}^{t-1}\frac{1}{\exp\left(h_{\tau+1}\right)^2}\mathbf{x}_\tau^{(n)}rx_{\tau+1}^{(n)}\right\}.\end{aligned}$$

The posterior distribution for $p\left(\sigma_\xi^{-2}\Big|\,\mu,\boldsymbol{\beta},h^t,\mathcal{D}^t\right)$ is readily available using

$$\sigma_\xi^{-2}\Big|\,\mu,\boldsymbol{\beta},h^t,\mathcal{D}^t\sim\mathcal{G}\left(\left[\frac{\underline{k}_\xi+\sum_{\tau=2}^{t-1}\left(h_{\tau+1}-h_\tau\right)^2}{t-1}\right]^{-1},t-1\right). \tag{A-18}$$

Finally, draws from the predictive density $p\left(rx_{t+1}^{(n)}\Big|\,\mathcal{D}^t\right)$ can be obtained by noting that

$$\begin{aligned}p\left(rx_{t+1}^{(n)}\Big|\,\mathcal{D}^t\right) &= \int_{\mu,\beta,h^{t+1},\sigma_\xi^{-2}}p\left(rx_{t+1}^{(n)}\Big|\,h_{t+1},\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right)\\ &\quad\times p\left(h_{t+1}|\,\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right)\\ &\quad\times p\left(\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2}\Big|\,\mathcal{D}^t\right)d\mu d\boldsymbol{\beta}dh^{t+1}d\sigma_\xi^{-2}.\end{aligned} \tag{A-19}$$

The first term in the integral above, $p\left(rx_{t+1}^{(n)}\Big|\,h_{t+1},\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right)$, represents the period $t+1$ predictive density of bond excess returns, treating model parameters as if they were known with certainty, and so is straightforward to calculate. The second term in the integral, $p\left(h_{t+1}|\,\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right)$, reflects how period $t+1$ volatility may drift away from h_t over time. Finally, the last term in the integral, $p\left(\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2}\Big|\,\mathcal{D}^t\right)$, measures parameter uncertainty in the sample.

To obtain draws for $p\left(rx_{t+1}^{(n)}\Big|\,\mathcal{D}^t\right)$, we proceed in three steps:

1. Simulate from $p\left(\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2}\Big|\,\mathcal{D}^t\right)$: draws from $p\left(\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2}\Big|\,\mathcal{D}^t\right)$ are obtained from the Gibbs sampling algorithm described above.

2. Simulate from $p\left(h_{t+1}|\,\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right)$: having processed data up to time t, the next step is to simulate the future volatility, h_{t+1}. For a given h_t and σ_ξ^{-2}, note that μ and β and the history of volatilities up to t become redundant, i.e., $p\left(h_{t+1}|\,\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right)=$

$p\left(h_{t+1}|\, h_t, \sigma_\xi^{-2}, \mathcal{D}^t\right)$. Note also that (24) along with the distributional assumptions made on $\xi_{\tau+1}$ imply that

$$h_{t+1}|\, h_t, \sigma_\xi^{-2}, \mathcal{D}^t \sim \mathcal{N}\left(h_t, \sigma_\xi^2\right). \tag{A-20}$$

3. Simulate from $p\left(rx_{t+1}^{(n)}\middle|\, h_{t+1}, \mu, \boldsymbol{\beta}, h^t, \sigma_\xi^{-2}, \mathcal{D}^t\right)$: For a given h_{t+1}, μ, and β, note that h^t and σ_ξ^{-2} become redundant, i.e., $p\left(rx_{t+1}^{(n)}\middle|\, h_{t+1}, \mu, \boldsymbol{\beta}, h^t, \sigma_\xi^{-2}, \mathcal{D}^t\right) = p\left(rx_{t+1}^{(n)}\middle|\, h_{t+1}, \mu, \boldsymbol{\beta}, \mathcal{D}^t\right)$. Then use the fact that

$$rx_{t+1}^{(n)}\middle|\, h_{t+1}, \mu, \boldsymbol{\beta}, \mathcal{D}^t \sim \mathcal{N}\left(\mu + \boldsymbol{\beta}'\mathbf{x}_\tau^{(n)}, \exp\left(h_{t+1}\right)\right). \tag{A-21}$$

A.3 Time-varying Parameter Model

In addition to specifying prior distributions and hyperparameters for $[\mu, \boldsymbol{\beta}]'$ and σ_ε^2, the TVP model in (25)-(26) requires eliciting a joint prior for the sequence of time-varying parameters $\boldsymbol{\theta}^t = \{\boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_t\}$ and its variance covariance matrix, $\mathbf{Q}$. For $[\mu, \boldsymbol{\beta}]'$ and σ_ε^2, we follow the same prior choices made for the linear model:

$$\begin{bmatrix} \mu \\ \boldsymbol{\beta} \end{bmatrix} \sim \mathcal{N}\left(\underline{b}, \underline{V}\right), \tag{A-22}$$

and

$$\sigma_\varepsilon^{-2} \sim \mathcal{G}\left(s_{rx,t}^{-2}, \underline{v}_0\left(t-1\right)\right). \tag{A-23}$$

Turning to $\boldsymbol{\theta}^t$ and $\mathbf{Q}$, we first write $p\left(\boldsymbol{\theta}^t, \mathbf{Q}\right) = p\left(\boldsymbol{\theta}^t\middle|\, \mathbf{Q}\right) p\left(\mathbf{Q}\right)$, and note that (26) along with the assumption that $\boldsymbol{\theta}_1 = \mathbf{0}$ implies

$$p\left(\boldsymbol{\theta}^t\middle|\, \mathbf{Q}\right) = \prod_{\tau=1}^{t-1} p\left(\boldsymbol{\theta}_{\tau+1}|\, \boldsymbol{\theta}_\tau, \mathbf{Q}\right), \tag{A-24}$$

with $\boldsymbol{\theta}_{\tau+1}|\, \boldsymbol{\theta}_\tau, \mathbf{Q} \sim \mathcal{N}\left(\boldsymbol{\theta}_\tau, \mathbf{Q}\right)$. Thus, to complete the prior elicitation for $p\left(\boldsymbol{\theta}^t, \mathbf{Q}\right)$ we only need to specify priors for $\mathbf{Q}$, which we choose to follow an Inverted Wishart distribution

$$\mathbf{Q} \sim \mathcal{IW}\left(\underline{\mathbf{Q}}, t-2\right), \tag{A-25}$$

with

$$\underline{\mathbf{Q}} = \underline{k}_Q\left(t-2\right)\left[s_{rx,t}^2\left(\sum_{\tau=1}^{t-1}\mathbf{x}_\tau^{(n)}\mathbf{x}_\tau^{(n)\prime}\right)^{-1}\right]. \tag{A-26}$$

The constant $\underline{k}_Q$ controls the degree of variation in the time-varying regression coefficients $\boldsymbol{\theta}_\tau$, where larger values of $\underline{k}_Q$ imply greater variation in $\boldsymbol{\theta}_\tau$.[31] We set $\underline{k}_Q = 0.01$ to limit the extent to which the parameters can change over time.

[31] This ensures that the scale of the Wishart distribution for $\mathbf{Q}$ is specified to be a fraction of the OLS estimates of the variance covariance matrix $s_{rx,t}^2\left(\sum_{\tau=1}^{t-1}\mathbf{x}_\tau^{(n)}\mathbf{x}_\tau^{(n)\prime}\right)^{-1}$ multiplied by the degrees of freedom, $t-2$, since for the inverted-Wishart distribution the scale matrix can be interpreted as the sum of squared residuals. This approach is consistent with the literature on TVP-VAR models; see, e.g., Cogley et al. (2005) and Primiceri (2005).

To obtain draws from the joint posterior distribution $p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q}\middle|\,\mathcal{D}^t\right)$ under the TVP model we use the Gibbs sampler to draw recursively from the following conditional distributions:

1. $p\left(\boldsymbol{\theta}^t\middle|\,\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\mathbf{Q},\mathcal{D}^t\right)$.
2. $p\left(\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2}\middle|\,\boldsymbol{\theta}^t,\mathbf{Q},\mathcal{D}^t\right)$.
3. $p\left(\mathbf{Q}\middle|\,\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\boldsymbol{\theta}^t,\mathcal{D}^t\right)$.

We simulate from each of these blocks as follows. Starting with θ^t, we focus on $p\left(\boldsymbol{\theta}^t\middle|\,\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\mathbf{Q},\mathcal{D}^t\right)$. Define $rx_{\tau+1}^{(n)*}=rx_{\tau+1}^{(n)}-\mu-\beta'\mathbf{x}_\tau^{(n)}$ and rewrite (25) as follows:

$$rx_{\tau+1}^{(n)*}=\mu_\tau-\boldsymbol{\beta}_\tau'\mathbf{x}_\tau^{(n)}+\varepsilon_{\tau+1} \tag{A-27}$$

Knowledge of μ and $\boldsymbol{\beta}$ makes $rx_{\tau+1}^{(n)*}$ observable, and reduces (25) to the measurement equation of a standard linear Gaussian state space model with homoskedastic errors. Thus, the sequence of time-varying parameters $\boldsymbol{\theta}^t$ can be drawn from (A-27) using the algorithm of Carter and Kohn (1994).

Moving on to $p\left(\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2}\middle|\,\boldsymbol{\theta}^t,\mathbf{Q},\mathcal{D}^t\right)$, conditional on $\boldsymbol{\theta}^t$ it is straightforward to draw $\mu,\boldsymbol{\beta}$, and σ_ε^{-2} by applying standard results. Specifically,

$$\left[\begin{array}{c}\mu\\ \boldsymbol{\beta}\end{array}\right]\Bigg|\,\sigma_\varepsilon^{-2},\boldsymbol{\theta}^t,\mathbf{Q},\mathcal{D}^t\sim\mathcal{N}\left(\overline{b},\overline{V}\right), \tag{A-28}$$

and

$$\sigma_\varepsilon^{-2}\middle|\,\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\mathcal{D}^t\sim\mathcal{G}\left(\overline{s}^{-2},\overline{v}\right), \tag{A-29}$$

where

$$\begin{aligned}\overline{V} &= \left[\underline{V}^{-1}+\sigma_\varepsilon^{-2}\sum_{\tau=1}^{t-1}\mathbf{x}_\tau^{(n)}\mathbf{x}_\tau^{(n)\prime}\right]^{-1},\\ \overline{b} &= \overline{V}\left[\underline{V}^{-1}\underline{b}+\sigma_\varepsilon^{-2}\sum_{\tau=1}^{t-1}\mathbf{x}_\tau^{(n)}\left(rx_{\tau+1}^{(n)}-\mu_\tau-\boldsymbol{\beta}_\tau'\mathbf{x}_\tau^{(n)}\right)\right],\end{aligned} \tag{A-30}$$

$$\overline{s}^2=\frac{\sum_{\tau=1}^{t-1}\left(rx_{\tau+1}^{(n)*}-\mu_\tau-\boldsymbol{\beta}_\tau'\mathbf{x}_\tau^{(n)}\right)^2+\left(s_{rx,t}^2\times\underline{v}_0\left(t-1\right)\right)}{\overline{v}}, \tag{A-31}$$

and $\overline{v}=\underline{v}_0+(t-1)$. As for $p\left(\mathbf{Q}\middle|\,\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\boldsymbol{\theta}^t,\mathcal{D}^t\right)$, we have that

$$\mathbf{Q}\middle|\,\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\boldsymbol{\theta}^t,\mathcal{D}^t\sim\mathcal{IW}\left(\overline{Q},2t-3\right), \tag{A-32}$$

where

$$\overline{\mathbf{Q}}=\underline{\mathbf{Q}}+\sum_{\tau=1}^{t-1}\left(\boldsymbol{\theta}_{\tau+1}-\boldsymbol{\theta}_\tau\right)\left(\boldsymbol{\theta}_{\tau+1}-\boldsymbol{\theta}_\tau\right)'. \tag{A-33}$$

Finally, draws from the predictive density $p\left(rx_{t+1}^{(n)}\middle|\,\mathcal{D}^t\right)$ can be obtained by noting than

$$\begin{aligned} p\left(rx_{t+1}^{(n)}\middle|\,\mathcal{D}^t\right) &= \int_{\mu,\boldsymbol{\beta},\boldsymbol{\theta}^{t+1}\mathbf{Q},\sigma_\varepsilon^{-2}} p\left(rx_{t+1}^{(n)}\middle|\,\boldsymbol{\theta}_{t+1},\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right) \\ &\quad\times p\left(\boldsymbol{\theta}_{t+1}|\,\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right) \\ &\quad\times p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2}\middle|\,\mathcal{D}^t\right) d\mu d\boldsymbol{\beta} d\boldsymbol{\theta}^t d\mathbf{Q} d\sigma_\varepsilon^{-2}. \end{aligned} \tag{A-34}$$

The first term in the integral above, $p\left(rx_{t+1}^{(n)}\middle|\,\boldsymbol{\theta}_{t+1},\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)$, represents the period $t+1$ predictive density of bond excess returns, treating model parameters as if they were known with certainty, and so is straightforward to calculate. The second term in the integral, $p\left(\boldsymbol{\theta}_{t+1}|\,\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)$, reflects that the regression parameters may drift away from $\boldsymbol{\theta}_t$ over time. Finally, the last term in the integral, $p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2}\middle|\,\mathcal{D}^t\right)$, measures parameter uncertainty.

To obtain draws for $p\left(rx_{t+1}^{(n)}\middle|\,\mathcal{D}^t\right)$, we proceed in three steps:

1. Simulate from $p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2}\middle|\,\mathcal{D}^t\right)$: draws from $p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2}\middle|\,\mathcal{D}^t\right)$ are obtained from the Gibbs sampling algorithm described above;

2. Simulate from $p\left(\boldsymbol{\theta}_{t+1}|\,\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)$: For a given θ_t and Q, note that μ, β, σ_ε^{-2}, and the history of regression parameters up to t become redundant, i.e., $p\left(\boldsymbol{\theta}_{t+1}|\,\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right) = p\left(\boldsymbol{\theta}_{t+1}|\,\boldsymbol{\theta}_t,\mathbf{Q},\mathcal{D}^t\right)$. Note also that (26), along with the distributional assumptions made with regards to $\eta_{\tau+1}$, imply that

$$\boldsymbol{\theta}_{t+1}|\,\boldsymbol{\theta}_t,\mathbf{Q},\mathcal{D}^t \sim \mathcal{N}\left(\boldsymbol{\theta}_t,\mathbf{Q}\right). \tag{A-35}$$

3. Simulate from $p\left(rx_{t+1}^{(n)}\middle|\,\boldsymbol{\theta}_{t+1},\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)$: For a given $\boldsymbol{\theta}_{t+1}$, μ, $\boldsymbol{\beta}$, and σ_ε^{-2}, $\boldsymbol{\theta}^t$ and $\mathbf{Q}$ become redundant so $p\left(rx_{t+1}^{(n)}\middle|\,\boldsymbol{\theta}_{t+1},\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right) = p\left(rx_{t+1}^{(n)}\middle|\,\boldsymbol{\theta}_{t+1},\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)$. Then use the fact that

$$rx_{t+1}^{(n)}\middle|\,\boldsymbol{\theta}_{t+1},\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\mathcal{D}^t \sim \mathcal{N}\left((\mu+\mu_t)+(\boldsymbol{\beta}+\boldsymbol{\beta}_t)'\,\mathbf{x}_\tau^{(n)},\sigma_\varepsilon^2\right). \tag{A-36}$$

A.4 Time-varying Parameter, Stochastic Volatility Model

Our priors for the TVP-SV model combine the earlier choices for the TVP and SV models, i.e., (A-22) and (A-23) for the regression parameters, (A-7) and (A-8) for the SV component, and (A-25) and (A-26) for the TVP component.

To obtain draws from the joint posterior distribution $p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\mathbf{Q},h^t,\sigma_\xi^{-2}\middle|\,\mathcal{D}^t\right)$ under the TVP-SV model, we use the Gibbs sampler to draw recursively from the following five conditional distributions:

1. $p\left(\boldsymbol{\theta}^t\middle|\,\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\mathbf{Q},\mathcal{D}^t\right)$.

2. $p\left(\mu, \boldsymbol{\beta}, \sigma_{\varepsilon}^{-2} \middle| \boldsymbol{\theta}^t, \mathbf{Q}, \mathcal{D}^t\right)$.

3. $p\left(h^t \middle| \mu, \boldsymbol{\beta}, \sigma_{\xi}^{-2}, \mathcal{D}^t\right)$.

4. $p\left(\mathbf{Q} \middle| \mu, \boldsymbol{\beta}, \sigma_{\varepsilon}^{-2}, \boldsymbol{\theta}^t, \mathcal{D}^t\right)$.

5. $p\left(\sigma_{\xi}^{-2} \middle| \mu, \boldsymbol{\beta}, h^t, \mathcal{D}^t\right)$.

With minor modifications, these steps are similar to the steps described in the TVP and SV sections above. Draws from the predictive density $p\left(rx_{t+1}^{(n)} \middle| \mathcal{D}^t\right)$ can be obtained from

$$\begin{aligned} p\left(rx_{t+1}^{(n)} \middle| \mathcal{D}^t\right) &= \int_{\mu,\boldsymbol{\beta},\boldsymbol{\theta}^{t+1}\mathbf{Q},h^{t+1},\sigma_{\xi}^{-2}} p\left(rx_{t+1}^{(n)} \middle| \boldsymbol{\theta}_{t+1}, h_{t+1}, \mu, \boldsymbol{\beta}, \boldsymbol{\theta}^t, \mathbf{Q}, h^t, \sigma_{\xi}^{-2}, \mathcal{D}^t\right) \\ &\quad \times p\left(\boldsymbol{\theta}_{t+1}, h_{t+1} \middle| \mu, \boldsymbol{\beta}, \boldsymbol{\theta}^t, \mathbf{Q}, h^t, \sigma_{\xi}^{-2}, \mathcal{D}^t\right) \\ &\quad \times p\left(\mu, \boldsymbol{\beta}, \boldsymbol{\theta}^t, \mathbf{Q}, h^t, \sigma_{\xi}^{-2} \middle| \mathcal{D}^t\right) d\mu d\boldsymbol{\beta} d\boldsymbol{\theta}^{t+1} d\mathbf{Q} dh^{t+1} d\sigma_{\xi}^{-2}. \end{aligned} \tag{A-37}$$

and following the steps described in the SV and TVP sections above.

References

Abrahams, M., T. Adrian, R. K. Crump, and E. Moench (2013). Decomposing real and nominal yield curves. Federal Reserve of New York Staff Report No. 570, New York, NY.

Altavilla, C., R. Giacomini, and R. Costantini (2014). Bond returns and market expectations. *Journal of Financial Econometrics*, forthcoming.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006). Volatility and correlation forecasting. Volume 1 of *Handbook of Economic Forecasting*, pp. 777 – 878. Elsevier.

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*(4), pp. 953–966.

Ang, A., S. Dong, and M. Piazzesi (2007). No-arbitrage taylor rules. NBER Working paper 13448.

Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics 50*(4), 745 – 787.

Bikbov, R. and M. Chernov (2010). No-arbitrage macroeconomic determinants of the yield curve. *Journal of Econometrics 159*(1), 166 – 182.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*(2), pp. 205–251.

Campbell, J. Y. and R. J. Shiller (1991). Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies 58*(3), 495–514.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies 21*(4), 1509–1531.

Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika 81*(3), pp. 541–553.

Clark, T. E. and M. McCracken (2011). Testing for unconditional predictive ability. In M. Clements and D. Hendry (Eds.), *Oxford Handbook of Economic Forecasting*. Oxford University Press: Oxford.

Clark, T. E. and F. Ravazzolo (2014). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, in press.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138*(1), 291 – 311.

Cochrane, J. H. and M. Piazzesi (2005). Bond risk premia. *American Economic Review 95*(1), 138–160.

Cogley, T., S. Morozov, and T. J. Sargent (2005). Bayesian fan charts for u.k. inflation: Forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control 29*(11), 1893 – 1925.

Cogley, T., G. E. Primiceri, and T. J. Sargent (2010). Inflation-gap persistence in the us. *American Economic Journal: Macroeconomics 2*, 43–69.

Cogley, T. and T. J. Sargent (2002). Evolving post-world war ii u.s. inflation dynamics. In B. Bernanke and K. Rogoff (Eds.), *NBER Macroeconomics Annual 2001*. Cambridge, U.S.: MIT Press.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*(1), 157 – 181.

Dewachter, H., L. Iania, and M. Lyrio (2014). Information in the yield curve: A macro-finance approach. *Journal of Applied Econometrics 29*(1), 42–64.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Duffee, G. R. (2011). Information in (and not in) the term structure. *Review of Financial Studies 24*(9), 2895–2934.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), pp. 603–615.

Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press.

Engsted, T., S. Moller, and M. Sander (2013). Bond return predictability in expansions and recessions. CREATES research paper 2013-13.

Fama, E. F. and R. R. Bliss (1987). The information in long-maturity forward rates. *The American Economic Review 77*(4), pp. 680–692.

Geweke, J. (2001). A note on some limitations of crra utility. *Economics Letters 71*(3), 341 – 345.

Geweke, J. and G. Amisano (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*(2), 216 – 230.

Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics 164*(1), 130 – 141.

Gurkaynak, R. S., B. Sack, and J. H. Wright (2007). The u.s. treasury yield curve: 1961 to the present. *Journal of Monetary Economics 54*(8), 2291 – 2304.

Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting 23*(1), 1 – 13.

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics 99*(3), 560 – 580.

Jagannathan, R. and T. Ma (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance 58*(4), 1651–1684.

Johannes, M., A. Korteweg, and N. Polson (2014, April). Sequential learning, predictive regressions, and optimal portfolio returns. *Journal of Finance 69*(2), 611–644.

Joslin, S., M. Priebsch, and K. J. Singleton (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, forthcoming.

Kandel, S. and R. F. Stambaugh (1996). On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance 51*(2), pp. 385–424.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies 65*(3), 361–393.

Koop, G. (2003). *Bayesian Econometrics*. John Wiley & Sons, Ltd.

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *Review of Financial Studies 22*(12), 5027–5067.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. *The Journal of Business 60*(4), pp. 473–489.

Paye, B. S. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance 13*(3), 274 – 315.

Pettenuzzo, D., A. Timmermann, and R. Valkanov (2013). Forecasting stock returns under economic constraints. *Journal of Financial Economics forthcoming*.

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies 72*(3), 821–852.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies 23*(2), 821–862.

Sarno, L., P. Schneider, and C. Wagner (2014). The economic value of predicting bond risk premia: Can anything beat the expectations hypothesis. Unpublished working paper, Cass Business School.

Sims, C. A. and T. Zha (2006). Were there regime switches in u.s. monetary policy? *American Economic Review 96*(1), 54–81.

Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*(2), 293 – 335.

Stock, J. H. and M. W. Watson (2006). Forecasting with many predictors. Volume 1 of *Handbook of Economic Forecasting*, pp. 515 – 554. Elsevier.

Svensson, L. E. O. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. IMF Working Paper No. 94/114.

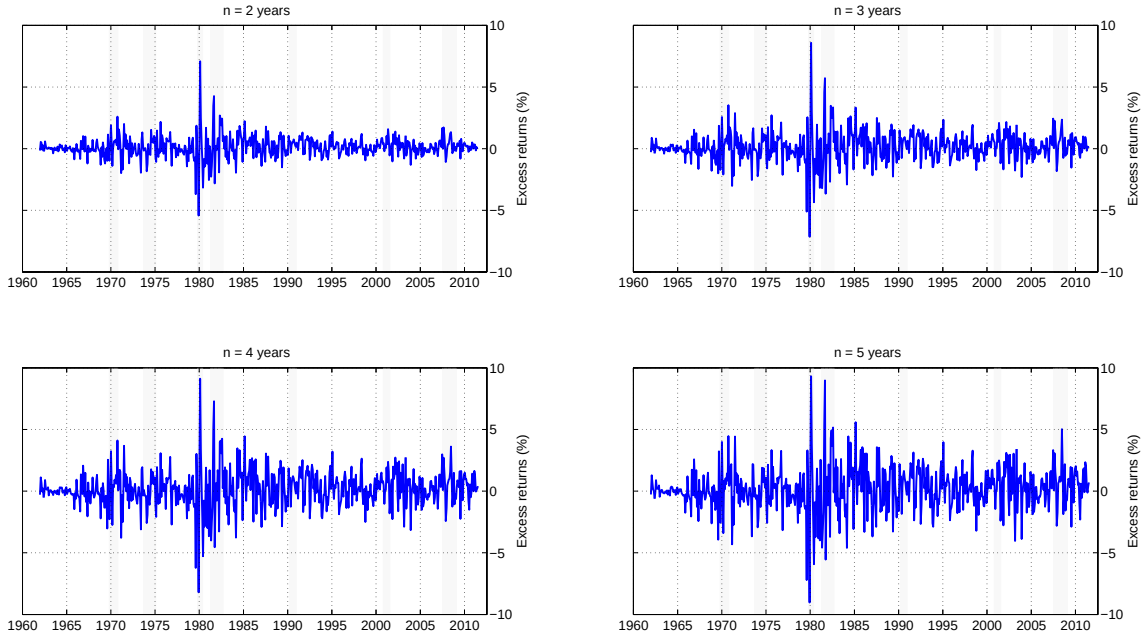
Thornton, D. L. and G. Valente (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *Review of Financial Studies 25*(10), 3141–3168.

Wei, M. and J. H. Wright (2013). Reverse regressions and long-horizon forecasting. *Journal of Applied Econometrics 28*(3), 353–371.

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

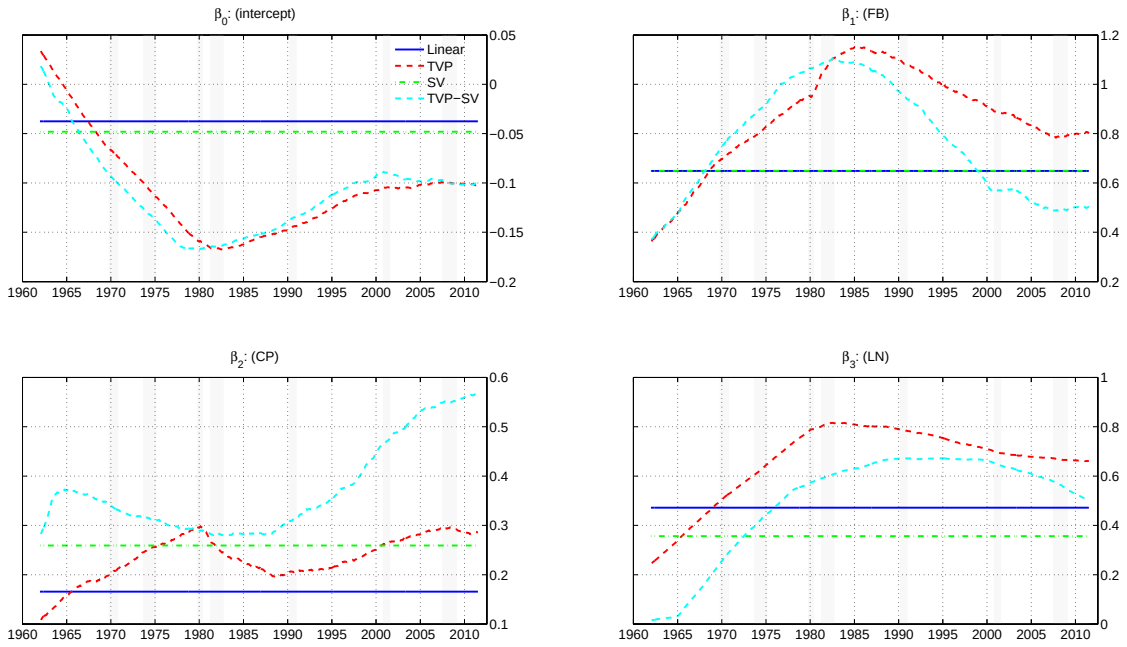
Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. *American Economic Review 101*(4), 1514–34.

Figure 1. **Bond excess returns**



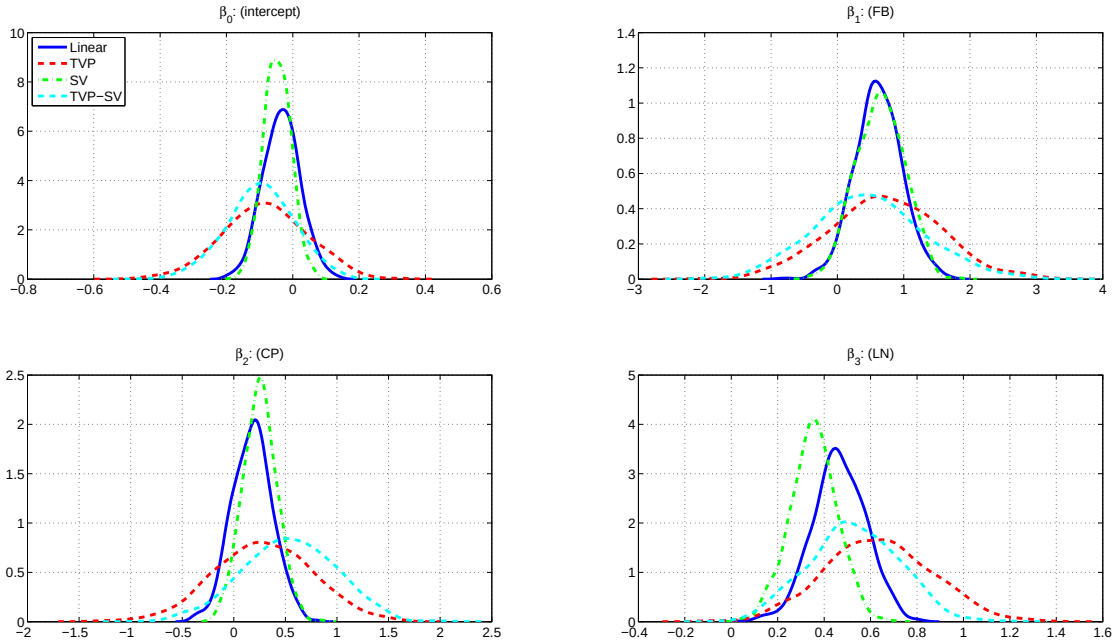
This figure shows time series of monthly bond excess returns (in percentage terms) for maturities (n) ranging from 2 years through 5 years. Monthly bond excess returns, $rx_{t+1/12}^{(n)}$, are computed from monthly yields, $y_t^{(n)}$, and are expressed in deviations from the 1-month T-bill rate, $rx_{t+1/12}^{(n)} = r_{t+1/12}^{(n)} - (1/12)y_t^{1/12}$, with $r_{t+1/12}^{(n)} = ny_t^{(n)} - (n - 1/12)y_{t+1/12}^{n-1/12}$. The sample ranges from January 1962 to December 2011.

Figure 2. **Parameter estimates for bond return forecasting model**



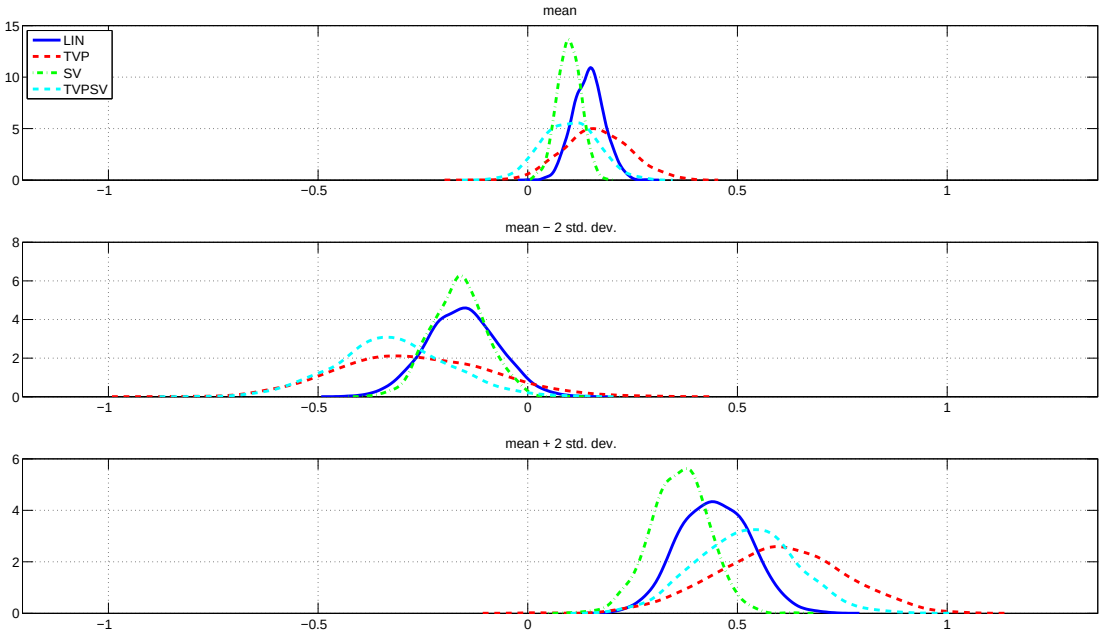
This figure displays parameter estimates for the FB-CP-LN model used to forecast monthly 3-year bond excess returns using as predictors the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) variables. The blue solid line represents the linear, constant coefficient model (Linear); the red dashed line tracks the parameter estimates for the time-varying parameter model (TVP); the green dashed-dotted line depicts the parameters for the stochastic volatility model (SV), while the dotted light-blue line shows estimates for the time-varying parameter, stochastic volatility (TVP-SV) model. The top left panel plots estimates of the intercept and the top right panel displays the coefficients on the FB predictor. The bottom left and right panels plot the coefficients on the CP and LN factors, respectively. The sample ranges from January 1962 to December 2011 and the parameter estimates are based on full-sample information.

Figure 3. **Posterior densities for model parameters**



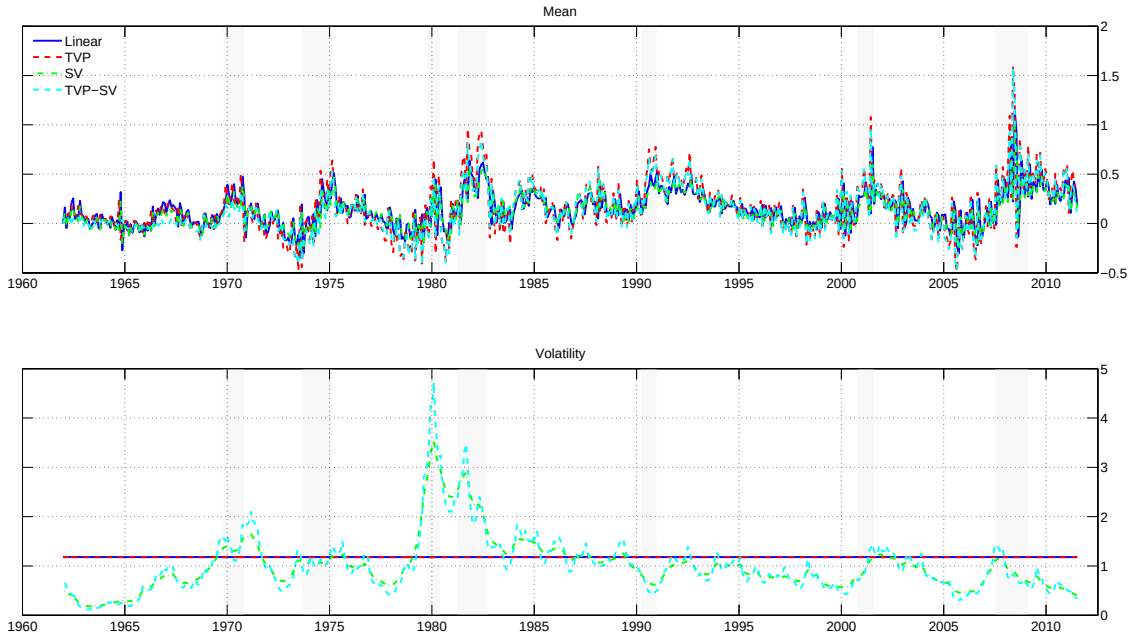
This figure displays posterior densities for the coefficients of the FB-CP-LN return model fitted to 3-year Treasury bonds, using as predictors the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) factors. The blue solid line represents the linear, constant coefficient (Linear) model; the red dashed line shows the parameter posterior density for the time-varying parameter (TVP) model; the green dashed-dotted line represents the stochastic volatility (SV) model, while the dotted light-blue line shows the posterior density for the time-varying parameter, stochastic volatility (TVP-SV) model. The first panel shows densities for the intercept. The second panel shows densities for the coefficient on the FB predictor. The third and fourth panels show densities for the coefficients on the CP and LN factors, respectively. The posterior density estimates shown here are based on their values as of 2011:12.

Figure 4. **Posterior densities for bond returns**



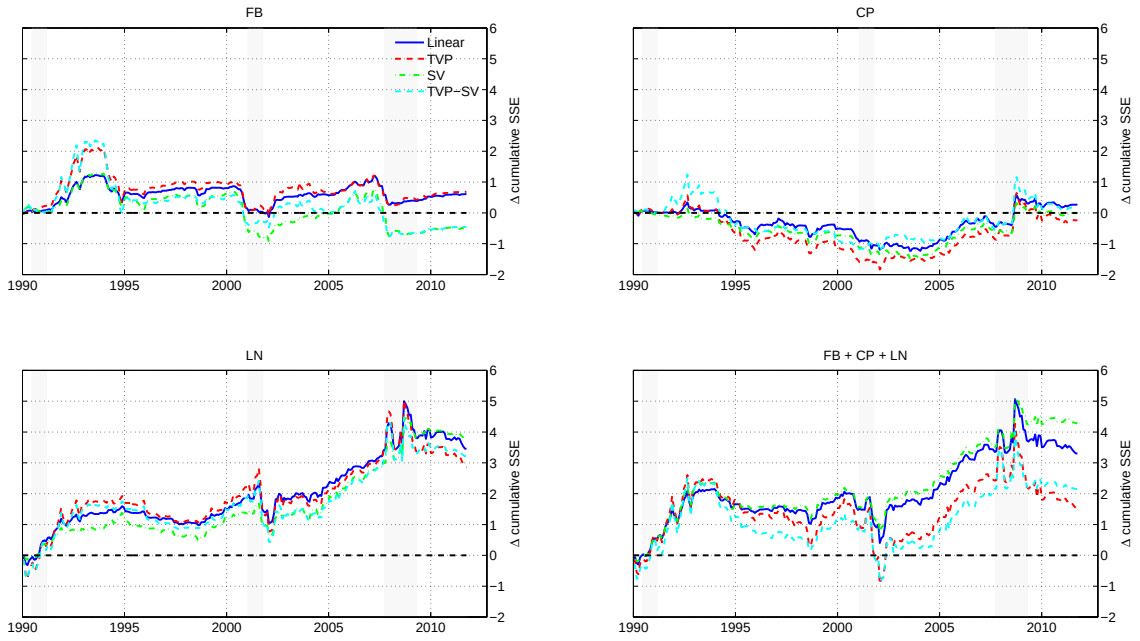
This figure shows the posterior density for excess returns on a three-year Treasury bond using the univariate Ludvigson-Ng (LN) state variable as a predictor. The LN variable is set at its sample mean $\overline{LN}$ (top panel), $\overline{LN} - 2stdev\,(LN)$ (middle panel), and $\overline{LN} + 2stdev\,(LN)$ (bottom panel). The blue solid line represents the linear, constant coefficient (Linear) model. the red dashed line tracks densities for the time-varying parameter (TVP) model. The green dashed-dotted line represents the stochastic volatility (SV) model, and the dotted light-blue line refers to the time varying parameter, stochastic volatility (TVP-SV) model. All posterior density estimates are based on the full data sample at the end of 2011.

Figure 5. **Conditional mean and volatility estimates for bond excess returns**



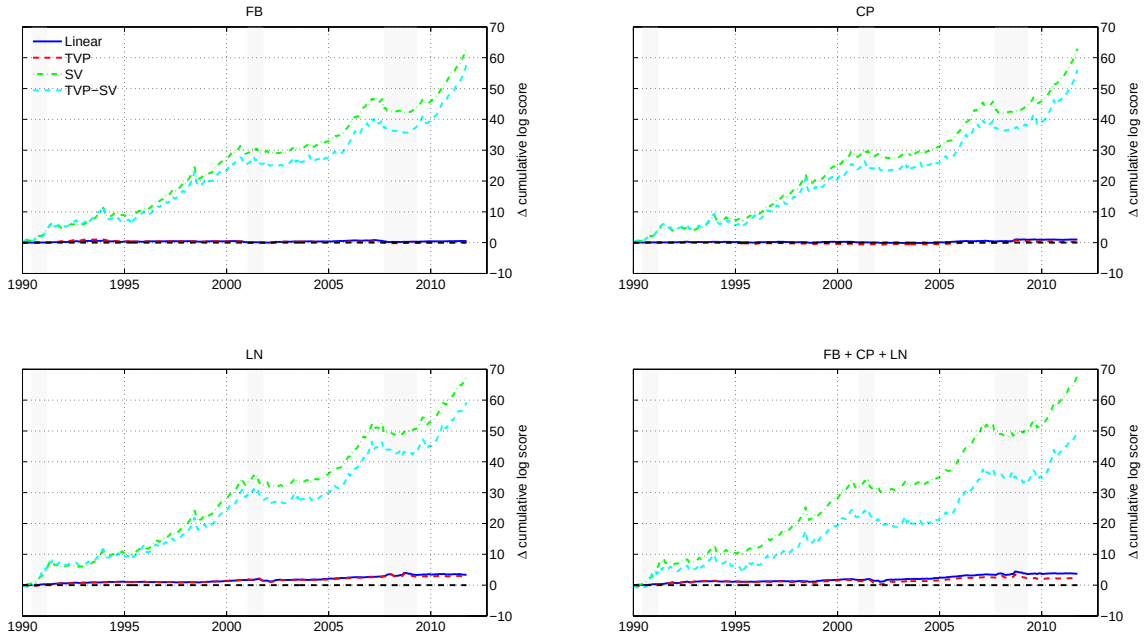
The top panel shows time-series of expected bond excess returns obtained from a range of models used to forecast monthly returns on a three-year Tresury bond using as predictors the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) factors. The blue solid line represents the linear, constant coefficient (Linear) model; the red dashed line tracks the time-varying parameter (TVP) model; the green dashed-dotted line depicts the stochastic volatility (SV) model, while the dotted light-blue line displays values for the time varying parameter, stochastic volatility (TVP-SV) model. The bottom panel displays volatility estimates for the FB-CP-LN models. The sample ranges from January 1962 to December 2011 and the estimates are based on full-sample information.

Figure 6. **Cumulative sum of squared forecast error differentials**



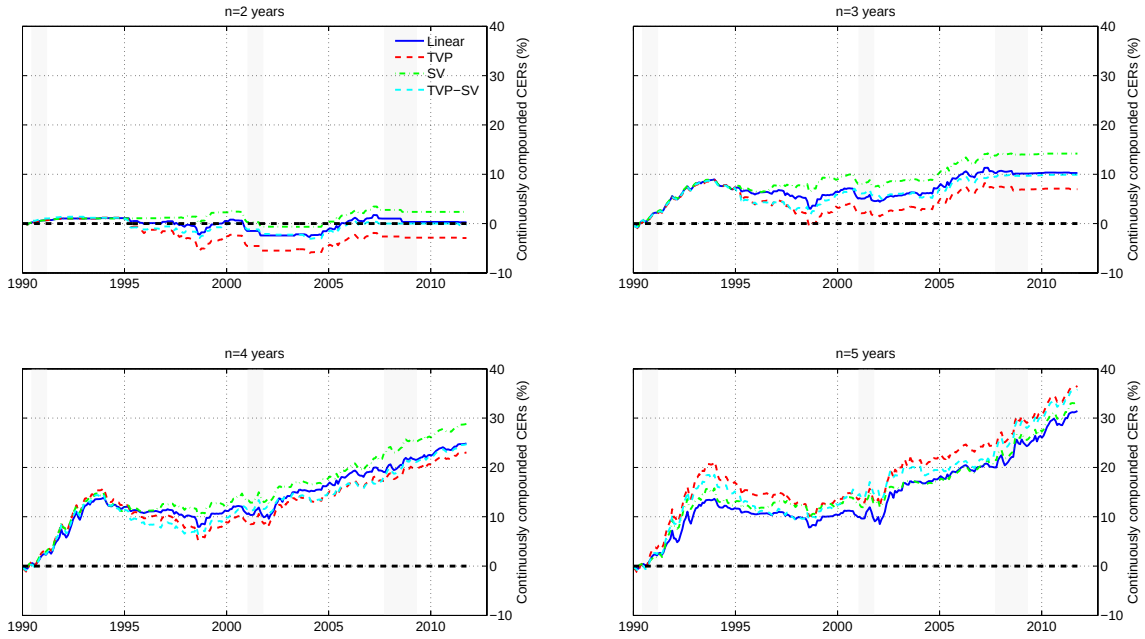
This figure shows the recursively calculated sum of squared forecast errors for the expectations hypothesis (EH) model minus the sum of squared forecast errors for a forecasting model with time-varying expected returns for a bond with a two year maturity, ($n = 2$). Each month we recursively estimate the parameters of the forecasting models and generate one-step-ahead forecasts of bond excess returns which are in turn used to compute out-of-sample forecasts. This procedure is applied to the EH model, which is our benchmark, as well as to forecasting models based on the Fama-Bliss (FB) predictor (1st window), the Cochrane-Piazzesi (CP) factor (2nd window), the Ludvigson-Ng (LN) factor (3rd window), and a multivariate model with all three predictors included (4th window). We then plot the cumulative sum of squared forecast errors (SSE_t) of the EH forecasts (SSE_t^{EH}) minus the corresponding value from the model with time-varying mean, $SSE_t^{EH} - SSE_t$. Values above zero indicate that a forecasting model with time-varying predictors produces more accurate forecasts than the EH benchmark, while negative values suggest the opposite. The blue solid line represents the linear, constant coefficient (Linear) model; the red dashed line tracks the time-varying parameter (TVP) model; the green dashed-dotted line represents the stochastic volatility (SV) model, while the dotted light-blue line refers to the time-varying parameter, stochastic volatility (TVP-SV) model. The out-of-sample period is 1990:01 - 2011:12.

Figure 7. **Cumulative sum of log-score differentials**



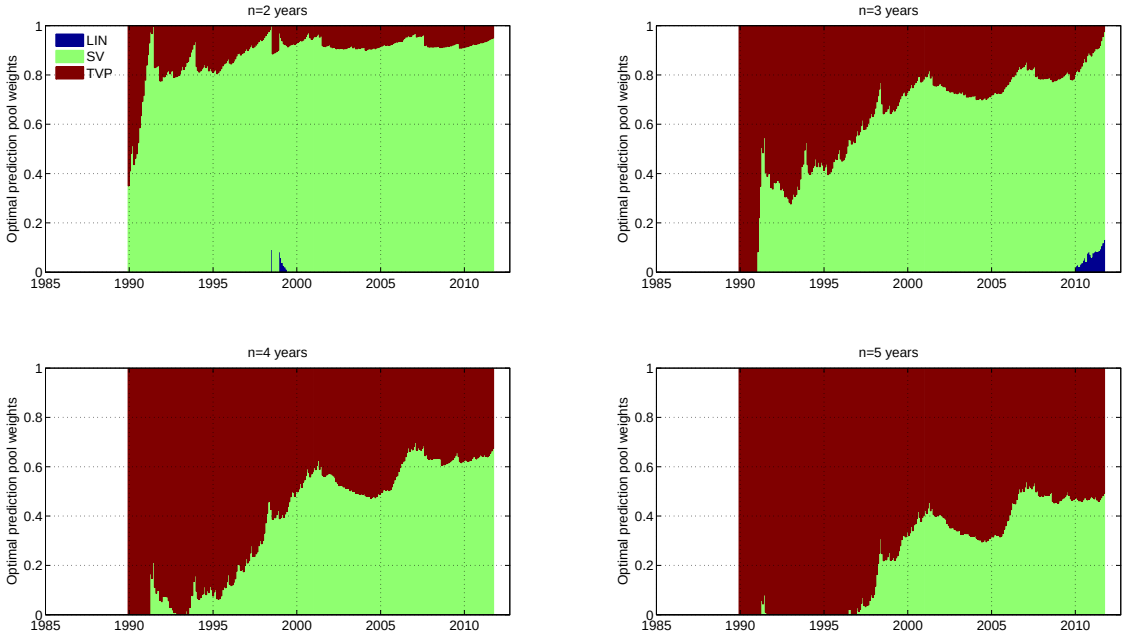
This figure shows the recursively calculated sum of log predictive scores from forecasting models with time-varying predictors minus the corresponding sum of log predictive scores for the EH model, using a 2-year Treasury bond. Each month we recursively estimate the parameters of the forecasting models and generate one-step-ahead density forecasts of bond excess returns which are in turn used to compute log-predictive scores. This procedure is applied to the benchmark EH model as well as to forecasting models based on the Fama-Bliss (FB) predictor (1st window), the Cochrane-Piazzesi (CP) factor (2nd window), the Ludvigson-Ng (LN) factor (3rd window), and a multivariate FB-CP-LN model (4th window). We then plot the cumulative sum of log predictive scores (LS_t) for the models with time-varying predictors minus the cumulative sum of log-predictive scores of the EH model, $LS_t - LS_t^{EH}$. Values above zero indicate that the time-varying mean model generates more accurate forecasts than the EH benchmark, while negative values suggest the opposite. The blue solid line represents the linear, constant coefficient (Linear) model; the red dashed line tracks the time-varying parameter (TVP) model; the green dashed-dotted line represents the stochastic volatility (SV) model, while the dotted light-blue line shows the time-varying parameter, stochastic volatility (TVP-SV) model. The out-of-sample period is 1990:01 - 2011:12.

Figure 8. **Economic value of out-of-sample bond return forecasts**



This figure plots cumulative certainty equivalent returns for the three-factor FB-CP-LN forecasting model that uses the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) factors as predictors, measured relative to the expectations hypothesis (EH) model. Each month we compute the optimal allocation to bonds and T-bills based on the predictive densities of bond excess returns. The investor is assumed to have power utility with a coefficient of relative risk aversion of ten and the weight on bonds is constrained to lie in the interval $[0, 0.99]$. Each panel displays a different bond maturity, ranging from 2 years (1st panel) to 5 years (4th panel) The blue solid line represents the linear, constant coefficient (Linear) model; the red dashed line tracks the time-varying parameter (TVP) model; the green dashed-dotted line represents the stochastic volatility (SV) model, while the dotted light-blue line shows results for the time-varying parameter, stochastic volatility (TVP-SV) model. The out-of-sample period is 1990:01 - 2011:12.

Figure 9. **Weights on different models in the optimal prediction pool**

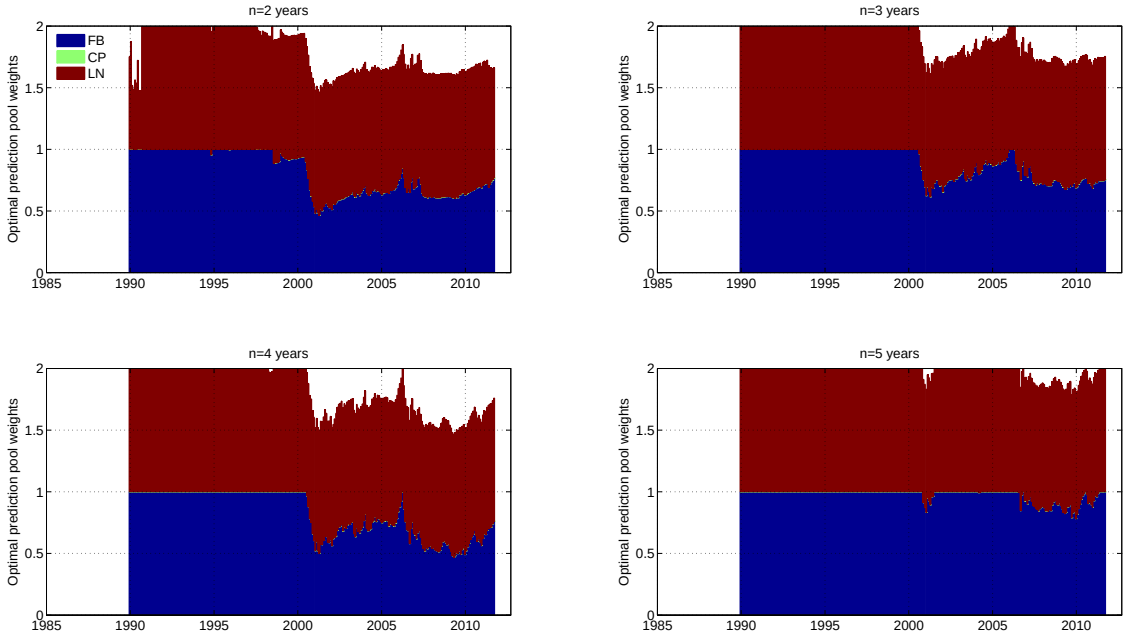


This figure plots the optimal weights on different model specifications in the predictive pool, computed in real time by solving the minimization problem

$$\mathbf{w}_t^* = \arg\max_{\mathbf{w}} \sum_{\tau=1}^{t-1} \log \left[\sum_{i=1}^{N} w_i \times S_{\tau+1,i} \right]$$

where $N = 21$ is the number of models considered and the solution is found subject to $\mathbf{w}_t^*$ belonging to the $N-$dimensional unit simplex. $S_{\tau+1,i}$ denotes the time $\tau + 1$ recursively computed log score for model i, $S_{\tau+1,i} = \exp(LS_{\tau+1,i})$. Blue bars show the weights on the linear (LIN) models in the optimal prediction pool, green bars show the weights assigned to the stochastic volatility (SV) models, and red bars show the weights on the time-varying parameter (TVP) models. Bond maturities range from 2 years (top left panel) to 5 years (bottom right panel).

Figure 10. **Weights on individual predictors in the optimal prediction pool**

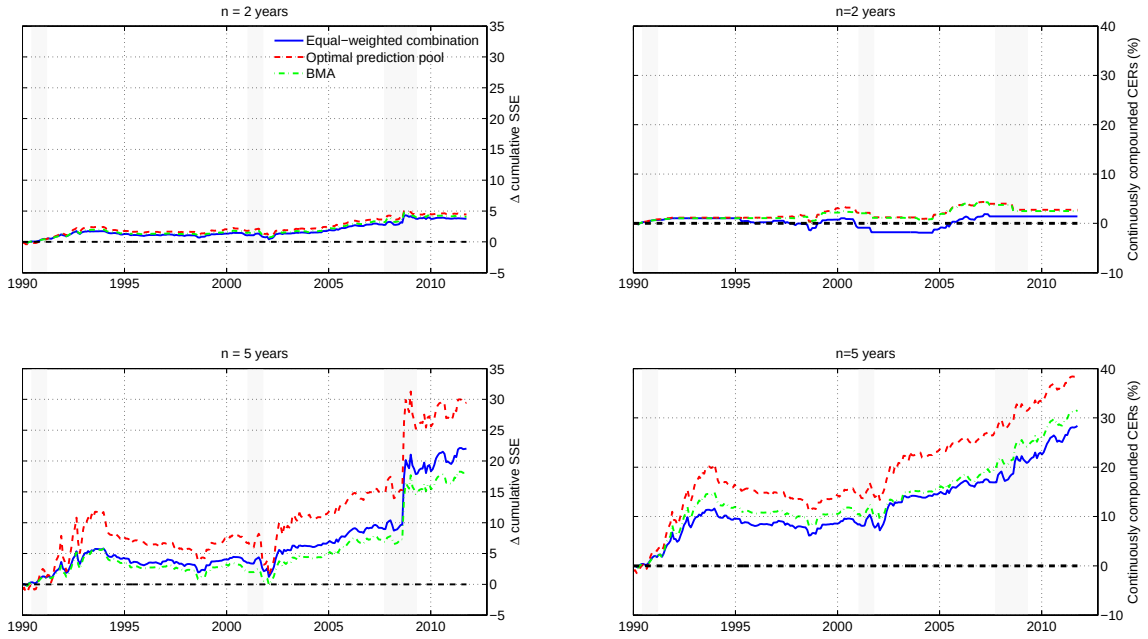


This figure plots the total weights on the individual predictors in the optimal predictive pool. At each point in time t, the weights are computed as $\mathcal{A}'\mathbf{w}_t$, where $\mathcal{A}$ is a 7×3 matrix representing all forecasting models by their unique combinations of zeros and ones and $\mathbf{w}_t$ is a 7×1 vector of period t optimal weights in the predictive pool, obtained in real time by solving the minimization problem

$$\mathbf{w}_t^* = \arg\max_{\mathbf{w}} \sum_{\tau=1}^{t-1} \log \left[\sum_{i=1}^{N} w_i \times S_{\tau+1,i} \right]$$

subject to $\mathbf{w}_t^*$ belonging to the $N-$dimensional unit simplex and N is the number of forecasting models. $S_{\tau+1,i}$ denotes the time $\tau+1$ log score for model i, $S_{\tau+1,i} = \exp(LS_{\tau+1,i})$. Blue bars show the combination weights associated with the Fama-Bliss (FB) factor; green bars show the weight assigned to the Cochrane-Piazzesi (CP) factor, and red bars show the weights assigned to the Ludvigson-Ng (LN) factor. Bond maturities range from 2 years (top left panel) to 5 years (bottom right panel).

Figure 11. **Out-of-sample forecasting performance for model combinations**



The left panels in this figure show the recursively calculated sum of squared forecast errors for the expectations hypothesis (EH) model minus the sum of squared forecast errors for three alternative forecasting models obtained using model combinations for two- (top panel) and five-year (bottom panel) bond maturities. Each month we recursively estimate the parameters of all forecasting models and generate one-step-ahead forecasts of bond excess returns which are in turn used to compute out-of-sample forecasts. This procedure is applied to the EH model, which is our benchmark, as well as to all forecasting models entering the model combinations. We then plot the cumulative sum of squared forecast errors (SSE_t) of the EH forecasts (SSE_t^{EH}) relative to the model combination forecasts, $SSE_t^{EH} - SSE_t^{COMB}$. Values above zero indicate that a model combination generates better performance than the EH benchmark, while negative values suggest the opposite. Three forecast combination schemes are considered, namely a simple equal-weighted combination, the optimal prediction pool of Geweke and Amisano (2011) and Bayesian Model Averaging (BMA) weights. The right panels plot the cumulative certainty equivalent returns of the same three combination schemes measured relative to the EH model. Each month we compute the optimal allocation to bonds and T-bills based on the predictive density of bond excess returns. The investor is assumed to have power utility with a coefficient of relative risk aversion of ten and the weight on bonds is constrained to lie in the interval $[0, 0.99]$. The top right panel displays results for a 2-year maturity while the bottom panel shows result for a 5-year maturity. The blue solid line represents the equal-weighted model combination, the red dotted line tracks the optimal prediction pool, and the green dashed-dotted line depicts results for the BMA combination.

Table 1. **Data Summary Statistics**

Panel A: Excess Returns					
	Bonds				Stocks
	2 years	3 years	4 years	5 years	*S&P*
mean	1.4147	1.7316	1.9868	2.1941	3.7327
st.dev.	2.9711	4.1555	5.2174	6.2252	15.2939
skew	0.4995	0.2079	0.0566	0.0149	-0.6314
kurt	14.8625	10.6482	7.9003	6.5797	5.3510
Sharpe	0.4761	0.4167	0.3808	0.3525	0.2441

Panel B: Predictors						
	Fama Bliss				CP	LN
	2-years	3-years	4-years	5-years		
mean	0.1078	0.1287	0.1451	0.1584	0.1533	0.1533
st.dev.	0.0996	0.1155	0.1278	0.1376	0.2126	0.3123
skew	-0.0716	-0.2379	-0.2234	-0.1434	0.7852	0.8604
kurt	3.7558	3.3782	3.0409	2.7548	5.2884	5.7283
AC(1)	0.8787	0.8995	0.9127	0.9227	0.6740	0.4109
AC(2)	0.7918	0.8211	0.8417	0.8580	0.5500	0.3537
AC(3)	0.7077	0.7493	0.7793	0.8030	0.6150	0.4690

Panel C: Correlation Matrix						
	FB-2	FB-3	FB-4	FB-5	CP	LN
FB-2	1.000	0.973	0.926	0.879	0.460	-0.087
FB-3		1.000	0.987	0.961	0.472	-0.049
FB-4			1.000	0.993	0.490	-0.010
FB-5				1.000	0.500	0.022
CP					1.000	0.184
LN						1.000

This table reports summary statistics for monthly bond excess returns and the predictor variables. Panel A reports the mean, standard deviation, skewness, kurtosis, and Sharpe ratio of bond excess returns for 2, 3, 4 and 5-year bond maturities (columns 1-4) as well as for excess returns on the S&P500 stock market index (last column). Excess returns are computed by subtracting the one-month T-bill rate. Means, standard deviations and Sharpe ratios are annualized. Panel B reports the same summary statistics for the predictors: the Fama-Bliss (FB) forward spreads (2, 3, 4, and 5 years), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) factors. Panel C reports the correlation matrix for the predictors. The sample period is 1962-2011.

Table 2. **Full-sample OLS estimates**

	FB	CP	LN	FB+CP	FB+LN	CP+LN	FB+CP+LN
				2 years			
β_{FB}	1.1648***			0.6621**	1.3592***		1.1221***
β_{CP}		0.6548***		0.5123***		0.4905***	0.2317
β_{LN}			0.6712***		0.7091***	0.6099***	0.6736***
R^2	0.0166	0.0246	0.0580	0.0277	0.0812	0.0708	0.0821
				3 years			
β_{FB}	1.3741***			0.7858**	1.4989***		1.2060***
β_{CP}		0.8784***		0.6769***		0.6561***	0.3299
β_{LN}			0.9068***		0.9342***	0.8248***	0.8876***
R^2	0.0158	0.0225	0.0540	0.0253	0.0732	0.0655	0.0742
				4 years			
β_{FB}	1.6661***			1.0071**	1.6936***		1.3499***
β_{CP}		1.1053***		0.8089***		0.8329***	0.4202
β_{LN}			1.1148***		1.1218***	1.0107***	1.0678***
R^2	0.0183	0.0226	0.0517	0.0266	0.0708	0.0635	0.0718
				5 years			
β_{FB}	1.9726***			1.2280**	1.9093***		1.4876***
β_{CP}		1.3612***		0.9640***		1.0441***	0.5502*
β_{LN}			1.3070***		1.2888***	1.1765***	1.2240***
R^2	0.0211	0.0242	0.0499	0.0292	0.0697	0.0630	0.0712

This table reports OLS estimates of the slope coefficients for seven linear models based on inclusion or exclusion of the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) predictor computed from a projection of the time series of cross-sectional averages of the 2, 3, 4, 5 bond excess returns on the 1, 2, 3, 4 and 5 year forward rates, and the Ludvigson-Ng (LN) predictor computed from a projection of the time-series of cross-sectional averages of the 2, 3, 4, 5 bond excess returns on five principal components obtained from a large panel of macroeconomic variables. Columns (1)-(3) report results for the univariate models, columns (4-6) for bivariate models and column (7) for the multivariate model that includes all three predictors. The last row in each panel reports the adjusted R^2. Stars indicate statistical significance based on Newey-West standard errors. ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 3. **Out-of-sample forecasting performance: R^2 values**

	Panel A: 2 years					Panel B: 3 years				
Model	OLS	LIN	SV	TVP	TVPSV	OLS	LIN	SV	TVP	TVPSV
FB	**0.09%***	**0.81%***	-0.65%	**0.91%***	-0.60%	**1.82%****	**1.42%****	**0.64%***	**2.13%****	**0.99%****
CP	-1.72%	**0.36%**	-0.01%	-0.32%	**0.24%***	-0.40%	**0.83%***	**0.24%**	**0.70%***	**0.56%***
LN	-0.40%	**4.59%*****	**5.05%*****	**3.80%*****	**4.29%*****	**2.65%*****	**4.66%*****	**3.80%*****	**4.10%*****	**3.90%*****
$FB+CP$	-1.34%	**0.86%***	**0.69%***	-0.14%	**0.45%****	**0.57%****	**1.39%****	**1.39%****	**1.40%****	**1.17%****
$FB+LN$	-1.70%	**5.32%*****	**5.94%*****	**3.09%*****	**3.86%*****	**1.36%*****	**5.62%*****	**5.28%*****	**4.44%*****	**3.78%*****
$CP+LN$	-3.49%	**3.57%*****	**4.70%*****	**2.01%*****	**2.70%*****	**0.72%*****	**4.11%*****	**4.05%*****	**3.25%*****	**2.76%*****
$FB+CP+LN$	-3.20%	**4.40%*****	**5.71%*****	**2.05%*****	**2.87%*****	**0.47%*****	**4.95%*****	**4.94%*****	**3.45%*****	**2.54%*****

	Panel C: 4 years					Panel D: 5 years				
Model	OLS	LIN	SV	TVP	TVPSV	OLS	LIN	SV	TVP	TVPSV
FB	**2.51%*****	**1.77%*****	**1.28%****	**2.54%*****	**1.83%****	**2.76%*****	**1.79%*****	**1.49%****	**2.81%*****	**1.95%****
CP	**0.37%**	**1.00%***	**0.32%**	**1.36%***	**1.13%***	**0.89%***	**0.91%***	**0.47%**	**1.69%***	**1.50%***
LN	**3.89%*****	**4.15%*****	**2.95%****	**4.54%*****	**4.08%*****	**4.47%*****	**3.53%*****	**2.30%****	**4.67%*****	**4.03%****
$FB+CP$	**1.64%****	**1.79%****	**1.66%****	**2.05%****	**1.74%****	**2.28%****	**1.96%****	**1.85%****	**2.55%****	**2.20%****
$FB+LN$	**2.46%*****	**5.19%*****	**4.69%*****	**4.67%*****	**4.07%*****	**2.79%*****	**4.91%*****	**4.05%*****	**5.09%*****	**4.21%*****
$CP+LN$	**2.57%*****	**3.92%*****	**3.45%****	**4.00%*****	**3.16%*****	**3.55%****	**3.61%****	**2.99%****	**4.24%****	**3.73%****
$FB+CP+LN$	**2.01%*****	**4.92%*****	**4.57%*****	**4.18%*****	**3.21%*****	**2.66%*****	**4.55%*****	**4.05%*****	**4.60%*****	**3.71%*****

This table reports out-of-sample R^2 values for seven prediction models based on the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) predictors fitted to monthly bond excess returns, rx_{t+1}, measured relative to the one-month T-bill rate. The R^2_{OoS} is measured relative to the EH model: $R^2_{OoS} = 1 - \frac{\sum_{\tau=\underline{t}-1}^{\overline{t}-1}(r_{xt+1}-\hat{r}_{xt+1|t})^2}{\sum_{\tau=\underline{t}-1}^{\overline{t}-1}(r_{xt+1}-\bar{r}_{xt+1|t})^2}$ where $\hat{r}_{xt+1|t}$ is the conditional mean of bond returns based on a regression of monthly excess returns on an intercept and lagged predictor variable(s), x_t: $r_{xt+1} = \mu + \beta' x_t + \varepsilon_{t+1}$. $\bar{r}_{t+1|t}$ is the forecast from the EH model which assumes that the βs are zero. We report results for five specifications: (i) ordinary least squares (OLS), (ii) a linear specification with constant coefficients and constant volatility (LIN), (iii) a model that allows for stochastic volatility (SV), (iv) a model that allows for time-varying coefficients (TVP) and (v) a model that allows for both time-varying coefficients and stochastic volatility ($TVPSV$). The out-of-sample period starts in January 1990 and ends in December 2011. We measure statistical significance relative to the expectation hypothesis model using the Clark and West (2007) test statistic. * significance at 10% level; ** significance at 5% level; *** significance at 1% level.

Table 4. **Out-of-sample forecasting performance: predictive likelihood**

	Panel A: 2 years				Panel B: 3 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	**0.002**	**0.248*****	-0.001	**0.220*****	**0.003**	**0.121*****	**0.001**	**0.097*****
CP	**0.004**	**0.239*****	-0.000	**0.202*****	**0.005****	**0.119*****	**0.002**	**0.095****
LN	**0.009***	**0.252*****	**0.005**	**0.207*****	**0.012****	**0.130*****	**0.005**	**0.087****
$FB+CP$	**0.004**	**0.243*****	-0.000	**0.194*****	**0.005***	**0.120*****	**0.002**	**0.083****
$FB+LN$	**0.014****	**0.257*****	**0.007**	**0.204*****	**0.015****	**0.131*****	**0.009**	**0.075****
$CP+LN$	**0.010***	**0.240*****	**0.004**	**0.183*****	**0.012****	**0.126*****	**0.007**	**0.073***
$FB+CP+LN$	**0.011****	**0.250*****	**0.004**	**0.176*****	**0.014****	**0.125*****	**0.006**	**0.047**
	Panel C: 4 years				Panel D: 5 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	**0.006****	**0.067*****	**0.004**	**0.045***	**0.006****	**0.033***	**0.006**	**0.016**
CP	**0.006****	**0.066*****	**0.005**	**0.046***	**0.005****	**0.028**	**0.007**	**0.015**
LN	**0.013****	**0.076*****	**0.011**	**0.041**	**0.012****	**0.036***	**0.013**	**0.013**
$FB+CP$	**0.008****	**0.063****	**0.006**	**0.034**	**0.008****	**0.033***	**0.009**	**0.004**
$FB+LN$	**0.016****	**0.075*****	**0.012**	**0.026**	**0.016****	**0.040***	**0.014**	-0.002
$CP+LN$	**0.013****	**0.072****	**0.012**	**0.026**	**0.014****	**0.037***	**0.014**	-0.004
$FB+CP+LN$	**0.016****	**0.072****	**0.012**	**0.002**	**0.017****	**0.033**	**0.014**	-0.020

This table reports the log predictive score for seven forecasting models that allow for time-varying predictors relative to the log-predictive score computed under the expectation hypothesis (EH) model. The seven forecasting models use the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) combination of forward rates, the Ludvigson-Ng (LN) macro factor, and combinations of these. Positive values of the test statistic indicate that the model with time-varying predictors generates more precise forecasts than the EH benchmark. We report results for a linear specification with constant coefficients and constant volatility (LIN), a model that allows for stochastic volatility (SV), a model that allows for time-varying coefficients (TVP) and a model that allows for both time-varying coefficients and stochastic volatility ($TVPSV$). The results are based on out-of-sample estimates over the sample period 1990 - 2011. ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level.

Table 5. **Out-of-sample economic performance of bond portfolios**

Constrained Weights

	Panel A: 2 years				Panel B: 3 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-0.23%	-0.25%	-0.21%	-0.25%	**0.05%**	**0.17%**	**0.18%**	**0.08%**
CP	-0.19%	**0.04%**	-0.24%	**0.06%**	-0.08%	**0.21%**	-0.16%	**0.28%***
LN	**0.11%**	**0.10%**	**0.09%**	**0.09%**	**0.57%*****	**0.64%*****	**0.66%*****	**0.66%*****
$FB+CP$	-0.21%	-0.13%	-0.27%	-0.10%	**0.01%**	**0.26%**	**0.08%**	**0.16%**
$FB+LN$	**0.07%**	**0.12%**	-0.07%	**0.01%**	**0.53%****	**0.67%*****	**0.38%***	**0.42%***
$CP+LN$	**0.09%**	**0.19%****	-0.04%	**0.01%**	**0.49%****	**0.67%*****	**0.46%***	**0.48%****
$FB+CP+LN$	**0.05%**	**0.14%**	-0.09%	**0.02%**	**0.47%****	**0.62%*****	**0.32%**	**0.43%****

	Panel C: 4 years				Panel D: 5 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	**0.46%****	**0.58%***	**0.56%****	**0.63%****	**0.69%****	**0.74%****	**0.84%****	**0.88%****
CP	**0.12%**	**0.25%**	**0.15%**	**0.28%**	**0.23%**	**0.35%**	**0.31%**	**0.43%**
LN	**0.90%*****	**0.96%*****	**1.19%*****	**1.22%*****	**0.90%*****	**0.97%*****	**1.37%*****	**1.47%*****
$FB+CP$	**0.38%***	**0.61%****	**0.46%**	**0.62%****	**0.62%****	**0.81%****	**0.71%***	**0.79%***
$FB+LN$	**1.03%*****	**1.17%*****	**0.91%*****	**0.83%****	**1.25%*****	**1.33%*****	**1.26%*****	**1.19%****
$CP+LN$	**0.73%*****	**1.01%*****	**0.97%*****	**0.97%*****	**0.79%*****	**1.03%*****	**1.13%*****	**1.28%*****
$FB+CP+LN$	**0.98%*****	**1.13%*****	**0.84%****	**0.90%*****	**1.11%*****	**1.21%*****	**1.21%****	**1.19%****

Unconstrained Weights

	Panel E: 2 years				Panel F: 3 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	**0.21%**	**0.99%**	**0.59%***	-1.39%	**0.49%****	**0.85%**	**0.99%****	**0.09%**
CP	**0.23%**	**1.32%**	**0.47%**	-5.62%	**0.29%**	**0.91%**	**0.53%**	-0.86%
LN	**2.62%*****	**6.90%*****	**3.45%*****	**5.69%***	**2.05%*****	**3.89%*****	**2.86%*****	**3.32%**
$FB+CP$	**0.34%**	**1.40%**	**0.59%**	-6.34%	**0.51%***	**1.10%**	**0.84%***	-2.26%
$FB+LN$	**2.91%*****	**6.79%****	**3.58%*****	**7.06%****	**2.52%*****	**4.04%****	**3.40%*****	**3.13%**
$CP+LN$	**2.39%*****	**5.04%**	**2.86%*****	**1.32%**	**1.86%*****	**3.00%***	**2.50%****	**0.65%**
$FB+CP+LN$	**2.68%*****	**5.42%***	**3.26%*****	**4.83%***	**2.26%*****	**3.15%***	**2.96%*****	**1.65%**

	Panel G: 4 years				Panel H: 5 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	**0.67%****	**0.74%**	**1.16%****	**0.42%**	**0.73%****	**0.81%***	**1.27%****	**0.51%**
CP	**0.34%**	**0.50%**	**0.62%**	-0.45%	**0.32%**	**0.49%**	**0.67%**	-0.04%
LN	**1.69%*****	**2.41%*****	**2.46%****	**0.84%**	**1.39%*****	**1.69%*****	**2.08%****	**0.85%**
$FB+CP$	**0.67%****	**0.91%**	**0.97%***	-1.23%	**0.77%****	**0.88%**	**1.12%***	-0.76%
$FB+LN$	**2.19%*****	**2.86%****	**2.87%****	**0.15%**	**1.98%*****	**2.19%****	**2.48%****	-0.45%
$CP+LN$	**1.55%****	**2.29%****	**2.13%****	-0.26%	**1.33%****	**1.71%****	**1.73%***	-0.83%
$FB+CP+LN$	**2.03%*****	**2.68%****	**2.53%****	-1.12%	**1.79%*****	**1.92%****	**2.11%***	-0.60%

This table reports annualized certainty equivalent return values for portfolio decisions based on recursive out-of-sample forecasts of bond excess returns. Each period an investor with power utility and coefficient of relative risk aversion of 10 selects 2, 3, 4, or 5-year bond and 1-month T-bills based on the predictive density implied by a given model. The seven forecasting models use the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) combination of forward rates, the Ludvigson-Ng (LN) macro factor, and combinations of these. We report results for a linear specification with constant coefficients and constant volatility (LIN), a model that allows for stochastic volatility (SV), a model that allows for time-varying coefficients (TVP) and a model with both time varying coefficients and stochastic volatility ($TVPSV$). Statistical significance is based on a one-sided Diebold-Mariano test applied to the out-of-sample period 1990-2011. * significance at 10% level; ** significance at 5% level; *** significance at 1% level.

Table 6. **Bond return predictability in expansions and recessions**

Model	LIN		SV		TVP		TVPSV	
	Exp	Rec	Exp	Rec	Exp	Rec	Exp	Rec
				Panel A: 2 years				
FB	2.86%	0.48%	3.05%	0.11%	4.53%	1.62%	5.49%	-0.47%
CP	1.53%	**4.08%****	1.87%	**3.24%***	3.68%	**4.63%***	4.02%	3.53%*
LN	1.23%	**12.03%****	1.95%	**7.87%*****	2.05%	**18.04%****	3.94%	**11.54%****
$FB+CP$	2.62%	**3.72%***	3.08%	2.78%	5.94%	5.57%*	5.94%	2.39%
$FB+LN$	4.97%	**12.83%*****	5.57%	**8.39%****	6.40%	**19.53%*****	7.99%	**13.42%****
$CP+LN$	2.66%	**13.47%****	3.41%	**8.81%*****	4.63%	**19.67%****	5.68%	**13.97%****
$FB+CP+LN$	4.86%	**13.51%*****	5.56%	**8.63%****	7.86%	**20.82%*****	8.35%	**14.82%****
				Panel B: 3 years				
FB	2.85%	-0.01%	3.13%	-0.64%	3.99%	0.81%	4.70%	-1.09%
CP	1.12%	**4.58%****	1.10%	**4.58%****	2.67%	**4.91%****	2.79%	**4.36%***
LN	1.69%	**11.58%****	1.87%	**8.73%*****	2.37%	**15.90%****	3.67%	**10.05%****
$FB+CP$	2.31%	**3.82%***	2.51%	**3.40%**	4.45%	**5.14%***	4.18%	2.66%
$FB+LN$	4.99%	**11.70%*****	5.20%	**8.27%****	6.12%	**17.07%*****	7.35%	**11.05%***
$CP+LN$	2.68%	**13.50%****	2.97%	**10.47%*****	4.01%	**17.75%****	4.60%	**12.88%****
$FB+CP+LN$	4.68%	**12.86%*****	4.79%	**9.72%*****	6.90%	**18.52%*****	6.79%	**12.91%****
				Panel C: 4 years				
FB	2.99%	0.16%	3.17%	-0.29%	3.89%	0.81%	4.33%	-0.62%
CP	0.96%	**5.36%****	0.98%	**5.24%****	2.40%	**5.54%****	2.09%	**5.60%****
LN	1.77%	**11.87%****	1.76%	**8.77%*****	2.63%	**14.91%****	3.53%	**9.84%****
$FB+CP$	2.32%	**4.34%***	2.45%	**4.03%**	3.92%	**5.71%***	3.32%	**3.86%**
$FB+LN$	4.94%	**11.85%****	4.96%	**8.16%****	6.05%	**16.24%****	6.95%	**10.26%***
$CP+LN$	2.65%	**14.13%****	2.69%	**11.22%*****	3.87%	**17.01%****	3.85%	**13.13%****
$FB+CP+LN$	4.63%	**13.21%*****	4.36%	**10.20%*****	6.35%	**18.03%*****	5.72%	**12.62%****
				Panel D: 5 years				
FB	3.09%	0.72%	3.16%	0.51%	3.84%	1.18%	4.10%	0.34%
CP	0.97%	**6.19%****	1.05%	**5.74%****	2.30%	**6.63%****	1.92%	**6.67%***
LN	1.79%	**11.81%****	1.59%	**8.12%*****	2.78%	**13.93%****	3.38%	**9.67%****
$FB+CP$	2.41%	**5.22%***	2.57%	**4.61%***	3.67%	**6.65%***	2.90%	**5.75%**
$FB+LN$	4.88%	**12.07%****	4.69%	**7.90%****	5.89%	**15.96%****	6.58%	**10.39%***
$CP+LN$	2.64%	**14.77%****	2.53%	**11.07%*****	3.81%	**16.82%****	3.68%	**13.01%****
$FB+CP+LN$	4.55%	**13.68%****	4.10%	**9.76%****	6.07%	**18.33%****	5.36%	**13.83%****

This table reports the R^2 from regressions of bond excess returns on the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) combination of forward rates, the Ludvigson-Ng (LN) macro factor, and combinations of these. We report results separately for expansions (Exp) and recessions (Rec) as defined by the NBER recession index. Results are shown for a linear specification with constant coefficients and constant volatility (LIN), a model that allows for stochastic volatility (SV), a model that allows for time-varying coefficients (TVP) and a model that allows for both time-varying coefficients and stochastic volatility ($TVPSV$). The R^2 in expansions is computed as $R^2_{i,0} = 1 - \frac{e_{i,0}{}' e_{i,0}}{e_{EH,0}{}' e_{EH,0}}$ where $e_{i,0}$ and $e_{EH,0}$ denote the vectors of residuals of the alternative and the benchmark model, respectively, during expansions. Similarly, the R^2 in recessions only uses the vector of residuals in recessions: $R^2_{i,1} = 1 - \frac{e_{i,1}{}' e_{i,1}}{e_{EH,1}{}' e_{EH,1}}$. We test whether the R^2 is higher in recessions than in expansions using a bootstrap methodology. * significance at 10% level; ** significance at 5% level; *** significance at 1% level.

Table 7. **Sharpe ratios in expansions and recessions**

Model	LIN		SV		TVP		TVPSV	
	Exp	Rec	Exp	Rec	Exp	Rec	Exp	Rec
	Panel A: 2 years							
FB	0.47	0.39	0.45	0.29	0.50	0.44	0.45	0.29
CP	0.45	**0.71**	0.42	**0.52**	0.45	**0.75**	0.44	**0.56**
LN	0.33	**1.43**	0.23	**0.80**	0.33	**1.62**	0.25	**1.08**
$FB+CP$	0.46	**0.62**	0.48	0.45	0.45	**0.72**	0.48	**0.49**
$FB+LN$	0.35	**1.43**	0.36	**0.82**	0.34	**1.62**	0.36	**1.17**
$CP+LN$	0.34	**1.56**	0.31	**0.87**	0.32	**1.72**	0.34	**1.25**
$FB+CP+LN$	0.35	**1.48**	0.37	**0.84**	0.32	**1.73**	0.39	**1.25**
	Panel B: 3 years							
FB	0.40	0.31	0.40	0.25	0.44	0.35	0.45	0.28
CP	0.37	**0.61**	0.47	**0.62**	0.38	**0.65**	0.46	**0.64**
LN	0.26	**1.25**	0.20	**0.91**	0.27	**1.43**	0.23	**1.14**
$FB+CP$	0.39	**0.53**	0.49	**0.52**	0.40	**0.59**	0.53	**0.55**
$FB+LN$	0.30	**1.24**	0.35	**0.90**	0.29	**1.41**	0.41	**1.20**
$CP+LN$	0.28	**1.39**	0.37	**1.09**	0.27	**1.54**	0.40	**1.37**
$FB+CP+LN$	0.31	**1.32**	0.39	**1.02**	0.28	**1.50**	0.50	**1.37**
	Panel C: 4 years							
FB	0.35	0.26	0.38	0.24	0.40	0.32	0.46	0.30
CP	0.33	**0.57**	0.45	**0.63**	0.35	**0.61**	0.50	**0.76**
LN	0.21	**1.16**	0.17	**0.90**	0.24	**1.33**	0.24	**1.23**
$FB+CP$	0.35	**0.47**	0.49	**0.53**	0.37	**0.54**	0.60	**0.66**
$FB+LN$	0.26	**1.14**	0.35	**0.91**	0.26	**1.30**	0.45	**1.25**
$CP+LN$	0.23	**1.28**	0.40	**1.17**	0.25	**1.43**	0.49	**1.59**
$FB+CP+LN$	0.27	**1.21**	0.42	**1.07**	0.26	**1.38**	0.63	**1.53**
	Panel D: 5 years							
FB	0.31	0.23	0.38	0.26	0.36	0.28	0.49	0.34
CP	0.28	**0.52**	0.45	**0.63**	0.32	**0.58**	0.54	**0.84**
LN	0.18	**1.07**	0.14	**0.83**	0.22	**1.23**	0.23	**1.27**
$FB+CP$	0.32	**0.45**	0.51	**0.55**	0.34	**0.50**	0.66	**0.74**
$FB+LN$	0.24	**1.06**	0.36	**0.86**	0.24	**1.22**	0.46	**1.25**
$CP+LN$	0.21	**1.21**	0.41	**1.14**	0.22	**1.34**	0.53	**1.64**
$FB+CP+LN$	0.25	**1.13**	0.45	**1.01**	0.23	**1.30**	0.66	**1.58**

This table reports the annualized Sharpe ratio computed from conditional mean and conditional volatility estimates implied by regressions of bond excess returns on the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) combination of forward rates, the Ludvigson-Ng (LN) macro factor, and combinations of these. We report results separately for expansions (Exp) and recessions (Rec) as defined by the NBER recession index. Results are shown for a linear specification with constant coefficients and constant volatility (LIN), a model that allows for stochastic volatility (SV), a model that allows for time-varying coefficients (TVP) and a model that allows for both time-varying coefficients and stochastic volatility ($TVPSV$).

Table 8. **Correlations between expected bond excess returns and economic variables**

	Panel A: GDP				Panel B: Inflation			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.14	0.16	0.12	0.20*	-0.11	-0.09	-0.10	-0.11
CP	-0.34***	-0.32***	-0.36***	-0.29***	-0.30***	-0.34***	-0.33***	-0.44***
LN	-0.61***	-0.61***	-0.62***	-0.60***	-0.38***	-0.37***	-0.31***	-0.30***
$FB+CP$	-0.17	-0.07	-0.21**	-0.13	-0.25**	-0.23**	-0.29***	-0.38***
$FB+LN$	-0.46***	-0.39***	-0.45***	-0.46***	-0.34***	-0.30***	-0.25**	-0.23**
$CP+LN$	-0.57***	-0.58***	-0.60***	-0.61***	-0.38***	-0.38***	-0.34***	-0.32***
$FB+CP+LN$	-0.48***	-0.43***	-0.49***	-0.51***	-0.35***	-0.33***	-0.31***	-0.29***

	Panel C: GDP Uncertainty				Panel D: Inflation Uncertainty			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.30***	0.27**	0.26**	0.15	0.02	-0.02	0.03	-0.11
CP	0.45***	0.43***	0.45***	0.44***	0.39***	0.37***	0.40***	0.24**
LN	0.51***	0.52***	0.53***	0.51***	0.49***	0.46***	0.48***	0.37***
$FB+CP$	0.45***	0.39***	0.42***	0.33***	0.27***	0.19*	0.31***	0.15
$FB+LN$	0.57***	0.55***	0.57***	0.52***	0.43***	0.37***	0.41***	0.34***
$CP+LN$	0.54***	0.55***	0.57***	0.58***	0.49***	0.47***	0.49***	0.41***
$FB+CP+LN$	0.57***	0.56***	0.57***	0.53***	0.45***	0.40***	0.44***	0.37***

This table reports the contemporaneous correlations between out-of-sample forecasts of excess returns on a two-year Treasury bond and real GDP growth (Panel A), inflation (Panel B), real GDP growth uncertainty (Panel C) and inflation uncertainty (Panel D). Real GDP growth is computed as $\Delta log(GDP_{t+1})$ where GDP_{t+1} is the real gross domestic product (GDPMC1 Fred mnemonic). Inflation is computed as $\Delta log(CPI_{t+1})$ where CPI is the consumer price index for all urban consumers (CPIAUCSL Fred mnemonic). Real GDP growth uncertainty is the cross-sectional dispersion (the difference between the 75th percentile and the 25th percentile) for real GDP forecasts from the Philadelphia Fed Survey of Professional Forecasters. Inflation uncertainty is the cross-sectional dispersion (the difference between the 75th percentile and the 25th percentile) for CPI forecasts from the Philadelphia Fed Survey of Professional Forecasters. The bond return prediction models use the Fama-Bliss (FB) forward spread predictor, the Cochrane-Piazzesi (CP) combination of forward rates, the Ludvigson-Ng (LN) macro factor, and combinations of these. We report results for a linear specification with constant coefficients and constant volatility (LIN), a model that allows for stochastic volatility (SV), a model that allows for time-varying coefficients (TVP) and a model that allows for both time-varying coefficients and stochastic volatility ($TVPSV$). Finally, we test whether the correlation coefficients are statistically different from zero. All results are based on the out-of-sample period 1990-2011. * significance at 10% level; ** significance at 5% level; *** significance at 1% level.

Table 9. **Economic and statistical performance of forecast combinations**

Method	2 years	3 years	4 years	5 years
	Panel A: Out-of-sample R^2			
OW	**5.92%*****	**5.55%*****	**5.05%*****	**5.16%*****
EW	**4.99%*****	**4.39%*****	**4.16%*****	**3.85%*****
BMA	**5.42%*****	**4.36%*****	**3.43%*****	**3.17%*****
	Panel B: Predictive Likelihood			
OW	**0.25*****	**0.11*****	**0.05*****	**0.03*****
EW	**0.14*****	**0.08*****	**0.05*****	**0.04*****
BMA	**0.25*****	**0.12*****	**0.05****	**0.02**
	Panel C: CER (constrained weights)			
OW	**0.15%**	**0.49%****	**0.98%*****	**1.30%*****
EW	**0.10%**	**0.53%*****	**0.96%*****	**1.02%*****
BMA	**0.14%**	**0.63%*****	**0.92%*****	**1.14%*****
	Panel D: CER (unconstrained weights)			
OW	**6.89%*****	**3.72%*****	**2.73%*****	**2.46%****
EW	**2.43%*****	**2.00%*****	**1.77%*****	**1.55%****
BMA	**6.02%*****	**3.32%*****	**2.09%*****	**1.73%****

This table reports out-of-sample results for the optimal predictive pool (OW) of Geweke and Amisano (2011), an equal-weighted (EW) model combination scheme, and Bayesian Model Averaging (BMA) applied to 21 forecasting models that use different predictors and are estimated using linear, stochastic volatility or time varying parameter methods. In each case the models and combination weights are estimated recursively using only data up to the point of the forecast. The R^2 values in Panel A use the out-of-sample R^2 measure proposed by Campbell and Thompson (2008). The predictive likelihood in Panel B is the value of the test for equal accuracy of the predictive density log-scores proposed by Clark and Ravazzolo (2014). CER values in Panels C and D are the annualized certainty equivalent returns derived for an investor with power utility and a coefficient of relative risk aversion of 10 who uses the posterior predictive density implied by the forecast combination. The forecast evaluation sample is 1990:01-2011:12. * significance at 10% level; ** significance at 5% level; *** significance at 1% level.