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

April 11, 2017

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 volatility dynamics and unspanned macro factors. 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. Such gains in predictive accuracy translate into higher risk-adjusted portfolio returns after accounting for estimation error and model uncertainty. Consistent with models featuring unspanned macro factors, our forecasts of future bond excess returns are strongly negatively correlated with survey forecasts of short rates.

Key words: bond returns; yield curve; macro factors; stochastic volatility; time-varying parameters; unspanned macro risk factors.

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 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 provide 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. 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 investments based on a no-predictability expectations hypothesis (EH) model. Sarno et al. (2016) reach a similar conclusion.[2]

To address this puzzling contradiction between the statistical and economic evidence on bond return predictability, we adopt an empirical modeling strategy that accounts for time-varying parameters, stochastic volatility and parameter estimation error and, thus, shares many features with the approach pioneered by Johannes et al. (2014) to explore predictability of stock returns. There are good economic reasons for considering these model features. First, bond prices, and thus bond returns, are sensitive to monetary policy and inflation prospects, both of which are known to shift over time.[3] This suggests that it is important to adopt a framework that accounts for time varying parameters. 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.[4] 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 esti-

---

[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.

[2] For example, Sarno et al. (2016) write that "The model predicts excess returns with high regression R^2s and high forecast accuracy but cannot outperform the expectations hypothesis out-of-sample in terms of economic value, showing a general contrast between statistical and economic metrics of forecast evaluation."

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

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

mation 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, in common with Johannes et al. (2014) we adopt a 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. Even after observing 50 years of monthly observations, the coefficients of the return prediction models are surrounded by considerable uncertainty and so accounting for estimation error turns out to be important. Indeed, we find many cases with strong improvements in forecasting performance as a result of incorporating estimation error.[5]

Second, we allow for time varying (stochastic) volatility in the bond excess return model. Stochastic volatility models do not, in general, lead to notably improved point forecasts of bond returns but they produce far better density forecasts which, when used by a risk averse investor to form a bond portfolio, generate better economic performance. In addition to reducing portfolio risk during periods with unusually high levels of volatility, the stochastic volatility models imply that investors load more heavily on risky bonds during times with relatively low interest rate volatility such as during the 1990s.

Third, our analysis allows for time variation in the regression parameters. We find evidence that accounting for time varying parameters can lead to more accurate forecasts and, when added to a model that already accounts for stochastic volatility, also improves on economic performance.

Fourth, we generalize the setup to include a multivariate asset allocation exercise where the optimal allocation to multiple risky bonds with different maturities is jointly determined. This extension requires modelling the dynamics of bond returns (and the various predictors) in a VAR setting with multivariate stochastic volatility and so is not a trivial extension of the setup of Johannes et al. (2014).

Fifth, and finally, we address model uncertainty through forecast combination methods. Model uncertainty is important in our analysis which considers a variety of univariate and multivariate models as well as different model specifications. 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, so our combinations

---

[5]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.

accommodate shifts in the relative forecasting performance of different models. The model combination results are generally better than the results for the individual models and thus suggest that model uncertainty can be effectively addressed through combination methods.[6]

Our empirical analysis uses the daily treasury yield data from Gurkaynak et al. (2007) to construct monthly excess returns for bond maturities between two and five years over the period 1962-2015. While previous studies have focused on the annual holding period, focusing on the higher frequency affords several advantages. Most obviously, it expands the number of non-overlapping observations, a point of considerable importance given the impact of parameter estimation error. Moreover, it allows us to identify short-lived dynamics in both first and second moments of bond returns which are missed by models of annual returns. This is an important consideration during events such as the global financial crisis of 2007-09 and around turning points of the economic cycle.

We conduct our analysis in the context of a three-variable model that 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). Since forecasting studies have found that simpler models often do well in out-of-sample experiments, we also consider simpler univariate models.[7]

To assess the statistical evidence on bond return predictability, we use our models to generate out-of-sample forecasts over the period 1990-2015. Our return forecasts are based on recursively updated parameter estimates and use only historically available information, thus allowing us to assess how valuable the forecasts would have been to investors in real time. Compared to the benchmark EH model that assumes no return predictability, we find that many of the return predictability models generate significantly positive out-of-sample R^2 values. Moreover, 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. 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. Our empirical results suggest that incorporating stochastic volatility and unspanned macro factors is important to understanding the economic gains from bond return predictability.

There are several reasons why our findings differ from studies such as Thornton and Valente (2012) and Sarno et al. (2016) which argue that the statistical evidence on bond return

---

[6]Using an iterated combination approach, Lin et al. (2016) uncover statistical and economic predictability in corporate bond returns

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

predictability does not translate into out-of-sample economic gains. Allowing for stochastic volatility leads to notable gains in economic performance for many models.[8] The inclusion of a composite macro factor as a predictor of bond returns is another important feature that differentiates our analysis from these earlier studies. Our results on forecast combinations also emphasize the importance of accounting for model uncertainty and the ability to capture changes in the performance of individual prediction models.

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 expected to be key drivers of time varying bond risk premia. We find 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). Using data from survey expectations we find that, consistent with a risk-premium story, 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.

Our finding that the macro factor of Ludvigson and Ng (2009) possesses considerable predictive power over bond excess returns out-of-sample implies that information embedded in the yield curve does not subsume information contained in such macro variables. We address possible explanations of this finding, including the unspanned risk factor models of Joslin et al. (2014) and Duffee (2011) which suggest that macro variables move forecasts of future bond excess returns and forecasts of future short rates by the same magnitude but in opposite directions. We find support for this explanation as our bond excess return forecasts are strongly negatively correlated with survey forecasts of future short rates.

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. Section 6 analyzes economic sources of bond return predictability such as recession risk, time variations in inflation uncertainty, and the presence of unspanned risk factors. Section 7 presents model combination results and Section 8 concludes.

---

[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.

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. 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]

$$
\begin{aligned}
y_t^{(n)} &= \beta_{0t} + \beta_{1t}\frac{1 - \exp\left(-\frac{n}{\tau_1}\right)}{\frac{n}{\tau_1}} + \beta_{2t}\left[\frac{1 - \exp\left(-\frac{n}{\tau_1}\right)}{\frac{n}{\tau_1}} - \exp\left(-\frac{n}{\tau_1}\right)\right] \\
&\quad + \beta_{3t}\left[\frac{1 - \exp\left(-\frac{n}{\tau_2}\right)}{\frac{n}{\tau_2}} - \exp\left(-\frac{n}{\tau_2}\right)\right].
\end{aligned}
\tag{1}
$$

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]

$$
\begin{aligned}
r_{t+h/12}^{(n)} &= p_{t+h/12}^{(n-h/12)} - p_t^{(n)} = n y_t^{(n)} - (n - h/12) y_{t+h/12}^{(n-h/12)}, \tag{2} \\
rx_{t+h/12}^{(n)} &= r_{t+h/12}^{(n)} - y_t^{h/12}(h/12). \tag{3}
\end{aligned}
$$

Here $p_t^{(n)}$ is the logarithm of the time t price of a bond with n periods to maturity. Similarly,

---

[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. The coefficients of the yield curve are estimated using daily cross-sections and thus avoid introducing look-ahead biases in the estimated yields.

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

forward rates are computed as[12]

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

2.2 Data Summary

We focus our analysis on monthly bond excess returns over the period 1962:01-2015:12. 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. Panel A.1 in Table 1 presents summary statistics for the four monthly excess return series. Returns on the two shortest maturities are right-skewed and fat-tailed, more so than the longer maturities.

Because the data used in our study differ from datasets used in most existing studies, it is worth highlighting the main differences and showing how they affect our data. First, there is a difference in how bond yields and returns are constructed. Studies such as Cochrane and Piazzesi (2005), Ludvigson and Ng (2009), and Thornton and Valente (2012) use data constructed using the method proposed by Fama and Bliss (1987) which sequentially constructs yields on long-term bonds from a set of estimated daily forward rates (see their Appendix A for more details). As described above, the bond returns in our analysis are, instead, based on daily yields constructed by Gurkaynak et al. (2007). Although the two approaches are different, they generate almost identical yields and excess return series with time-series correlations ranging between 0.991 to 0.9998 across the four bond maturities. Thus, we conclude that this difference matters little to our analysis.

More important is our use of one-month (non-overlapping) returns data as compared to the 12-month overlapping returns data used in many existing studies. Panels A.2 and A.3 in Table 1 provide summary statistics on the more conventional overlapping 12-month returns constructed either from our monthly data (Panel A.2) or as in Cochrane and Piazzesi (2005) (Panel A.3), using the Fama-Bliss CRSP files. The two series have very similar means which in turn are lower than the mean excess return on the monthly series in Panel A.1 due to the lower mean of the risk-free rate (1-month T-bill) used in Panel A.1 compared to the mean of the 12-month T-bill rate used in Panels A.2 and A.3. Comparing the monthly series in Panel A.1 to the 12-month series in Panels A.2 and A.3, we see that the serial correlation is much stronger in the 12-month series due to the smoothing effect of using overlapping returns.

Using monthly non-overlapping bond returns offers important advantages over the 12-month overlapping returns data which have been the focus of most studies in the literature. 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

---

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

missed by models focusing on the annual holding period. Bond returns recorded at the annual horizon easily overlook important variations around turning points of the economic cycle.

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).

To motivate our use of these three predictor variables, note that the n-period bond yield is related to expected future short yields and expected future excess returns (Duffee, 2013):

$$y_t^{(n)} = \frac{1}{n} \left(\sum_{j=0}^{n-1} E[y_{t+j}^{(1)}|\boldsymbol{z}_t] \right) + \frac{1}{n} \left(\sum_{j=0}^{n-1} E_t[rx_{t+j+1}^{(n-j)}|\boldsymbol{z}_t] \right), \tag{5}$$

where $rx_{t+j+1}^{(n-j)}$ is the excess return in period $t+j+1$ on a bond with $n-j$ periods to maturity and $E[.|\boldsymbol{z}_t]$ denotes the conditional expectation given market information at time t, $\boldsymbol{z}_t$. Equation (5) suggests that current yields or, equivalently, forward spreads should have predictive power over future bond excess returns and so motivates our use of these variables in the excess return regressions.

The use of non-yield predictors is more contentious. In fact, if the vector of conditioning information variables, $\boldsymbol{z}_t$, is of sufficiently low dimension, we can invert (5) to get $\boldsymbol{z}_t = g(\boldsymbol{y}_t)$. In this case information in the current yield curve subsumes all other predictors of future excess returns and so macro variables should be irrelevant when added to the prediction model. The unspanned risk factor models of Joslin et al. (2014) and Duffee (2011) offer an explanation for why macro variables help predict bond excess returns over and above information contained in the yield curve. These models suggest that the effect of additional state variable on expected future short rates and expected future bond excess returns cancel out in Equation (5). Such cancellations imply that the additional state variables do not show up in bond yields although they can have predictive power over bond excess returns.

Our predictor variables are computed as follows. The Fama-Bliss (FB) forward spreads are given by

$$fs_t^{(n,h)} = f_t^{(n-h/12,n)} - y_t^{(h/12)}(h/12). \tag{6}$$

The Cochrane-Piazzesi (CP) factor is formed from a linear combination of forward rates

$$CP_t^h = \hat{\boldsymbol{\gamma}}^{h\prime} \boldsymbol{f}_t^{(\boldsymbol{n}-h/12,\boldsymbol{n})}, \tag{7}$$

where

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

8

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

$$\frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} = \gamma_0^h + \gamma_1^h f_t^{(1-1/12,1)} + \gamma_2^h f_t^{(2-1/12,2)} + \gamma_3^h f_t^{(3-1/12,3)} + \gamma_4^h f_t^{(4-1/12,4)} + \gamma_5^h f_t^{(5-1/12,5)} + \bar{\varepsilon}_{t+h/12}. \tag{8}$$

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

$$x_{i,t} = \kappa_i g_t + \epsilon_{i,t}, \tag{9}$$

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, $\widehat{\boldsymbol{G}}_t = [\hat{g}_{1,t}, \hat{g}_{1,t}^3, \hat{g}_{3,t}, \hat{g}_{4,t}, \hat{g}_{8,t}]$ to obtain the LN factor[13]

$$LN_t^h = \hat{\boldsymbol{\lambda}}^{h\prime} \widehat{\boldsymbol{G}}_t, \tag{10}$$

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

$$\frac{1}{4}\sum_{n=2}^{5} rx_{t+h/12}^{(n)} = \lambda_0^h + \lambda_1^h \hat{g}_{1,t} + \lambda_2^h \hat{g}_{1,t}^3 + \lambda_3^h \hat{g}_{3,t} + \lambda_4^h \hat{g}_{4,t} + \lambda_5^h \hat{g}_{8,t} + \bar{\eta}_{t+h/12}. \tag{11}$$

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 0.71 and 0.39, respectively.

Panel C shows that the Fama-Bliss spreads are positively 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 0.13.

3 Return Prediction Models and Estimation Methods

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

---

[13]Ludvigson and Ng (2009) select this combination of factors using the Schwarz information criterion. To compute the LN factor, we use the FRED-MD dataset. This data was downloaded from https://research.stlouisfed.org/econ/mccracken/fred-databases/ and allows us to extend the original data of Ludvigson and Ng (2009) up to 2015. While not all variables are identical to those used in Ludvigsson and Ng, they are very similar and the corresponding principal components are very highly correlated. Before extracting the factors, each variable is transformed as described in the Appendix of McCracken and Ng (2015).

3.1 Model specifications

Our analysis considers the three predictor variables described in the previous section. Specifically, we consider three univariate models, each of which includes one of these three variables, along with a multivariate model that includes all three predictors for a total of four models:

1. Fama-Bliss (FB) univariate

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

2. Cochrane-Piazzesi (CP) univariate

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

3. Ludvigson-Ng (LN) univariate

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

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

$$rx_{t+h/12}^{(n)} = \beta_0 + \beta_1 fs_t^{(n,h)} + \beta_2 CP_t^h + \beta_3 LN_t^h + \varepsilon_{t+h/12}. \tag{15}$$

These models are in turn compared to the Expectation Hypothesis benchmark

$$rx_{t+h/12}^{(n)} = \beta_0 + \varepsilon_{t+h/12}, \tag{16}$$

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

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 produce better out-of-sample forecasts since their parameters may be imprecisely estimated.

To estimate the models we adopt a Bayesian approach that offers several advantages over the conventional estimation methods adopted by previous studies on 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 accomplished by our method which generates the (posterior) predictive return distribution. Third, our approach also allows us to handle model uncertainty (and model instability) by combining forecasting 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 depends on the data frequency.

3.2   Constant Coefficients and Constant Volatility

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

$$
\begin{aligned}
rx_{\tau+1}^{(n)} &= \mu + \boldsymbol{\beta}' \boldsymbol{x}_{\tau}^{(n)} + \varepsilon_{\tau+1}, \quad \tau = 1, ..., t-1, \\
\varepsilon_{\tau+1} &\sim \mathcal{N}(0, \sigma_{\varepsilon}^2).
\end{aligned}
\tag{17}
$$

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 for this (LIN) model. Following standard practice, the priors for the parameters μ and $\boldsymbol{\beta}$ in (17) are assumed to be normal and independent of σ_{ε}^2

$$
\left[\begin{array}{c} \mu \\ \boldsymbol{\beta} \end{array} \right] \sim \mathcal{N}\left(\underline{\boldsymbol{b}}, \underline{\boldsymbol{V}} \right),
\tag{18}
$$

where

$$
\underline{\boldsymbol{b}} = \left[\begin{array}{c} \overline{rx}_t^{(n)} \\ \boldsymbol{0} \end{array} \right], \quad \underline{\boldsymbol{V}} = \underline{\psi}^2 \left[\left(s_{rx,t}^{(n)} \right)^2 \left(\sum_{\tau=1}^{t-1} \boldsymbol{x}_{\tau}^{(n)} \boldsymbol{x}_{\tau}^{(n)\prime} \right)^{-1} \right],
\tag{19}
$$

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 $\underline{\boldsymbol{b}}$ 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. To avoid any look-ahead bias in the out-of-sample forecasting exercise, the prevailing mean is based only on information available at the time of the forecast which amounts to using the historical average at that point in time.

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 find that, at least for a reasonable range of values, the choice of priors has modest impact on our results. In (19), $\underline{\psi}$ 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 bond maturities for which fundamentals-based information is likely to be more important.

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

$$\sigma_\varepsilon^{-2} \sim \mathcal{G} \left(\left(s_{rx,t}^{(n)} \right)^{-2}, \underline{v}_0 \left(t - 1 \right) \right), \tag{20}$$

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}. 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

A large literature has found strong empirical evidence of time varying return volatility. We accommodate such effects through a simple stochastic volatility (SV) model:

$$rx_{\tau+1}^{(n)} = \mu + \boldsymbol{\beta}' \boldsymbol{x}_\tau^{(n)} + \exp\left(h_{\tau+1} \right) u_{\tau+1}, \tag{21}$$

where $h_{\tau+1}$ denotes the (log of) bond return volatility at time $\tau + 1$ and $u_{\tau+1} \sim \mathcal{N}(0, 1)$. The log-volatility $h_{\tau+1}$ is assumed to follow a stationary and mean reverting process:

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \tag{22}$$

where $\xi_{\tau+1} \sim \mathcal{N}\left(0, \sigma_\xi^2 \right)$, $|\lambda_1| < 1$, and u_τ and ξ_s are mutually independent for all τ and s. Appendix A explains how we estimate the SV model and set the priors.

3.4   Time varying Parameters

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 (17) to change over time:

$$\begin{aligned} rx_{\tau+1}^{(n)} &= \left(\mu + \mu_\tau \right) + \left(\boldsymbol{\beta} + \boldsymbol{\beta}_\tau \right)' \boldsymbol{x}_\tau^{(n)} + \varepsilon_{\tau+1}, \quad \tau = 1, ..., t - 1, \tag{23} \\ \varepsilon_{\tau+1} &\sim \mathcal{N}(0, \sigma_\varepsilon^2). \end{aligned}$$

The intercept and slope parameters $\boldsymbol{\theta}_\tau = \left(\mu_\tau, \boldsymbol{\beta}_\tau' \right)'$ are assumed to follow a zero-mean, stationary process

$$\boldsymbol{\theta}_{\tau+1} = diag\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}} \right) \boldsymbol{\theta}_\tau + \boldsymbol{\eta}_{\tau+1}, \tag{24}$$

12

where $\boldsymbol{\theta}_1 = \mathbf{0}$, $\boldsymbol{\eta}_{\tau+1} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right)$, and the elements in $\boldsymbol{\gamma_\theta}$ are restricted to lie between -1 and 1. In addition, ε_τ and $\boldsymbol{\eta}_s$ are mutually independent for all τ and s. The key parameter is $\boldsymbol{Q}$ 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. Again Appendix A provides details on how we estimate the model and set the priors.

3.5 Time varying Parameters and Stochastic Volatility

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

$$rx_{\tau+1}^{(n)} = (\mu + \mu_\tau) + (\boldsymbol{\beta} + \boldsymbol{\beta}_\tau)' \boldsymbol{x}_\tau^{(n)} + \exp\left(h_{\tau+1}\right) u_{\tau+1}, \tag{25}$$

with

$$\boldsymbol{\theta}_{\tau+1} = diag\left(\boldsymbol{\gamma_\theta}\right) \boldsymbol{\theta}_\tau + \boldsymbol{\eta}_{\tau+1}, \tag{26}$$

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

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \tag{27}$$

where $u_{\tau+1} \sim \mathcal{N}\left(0,1\right)$, $\boldsymbol{\eta}_{\tau+1} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{Q}\right)$, $\xi_{\tau+1} \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$ and u_τ, $\boldsymbol{\eta}_s$ and ξ_l are mutually independent for all τ, s, and l. Again we refer to Appendix A 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, we first report results based on full-sample estimates. This is followed by an out-of-sample analysis of the statistical evidence on return predictability.

4.1 Full-sample Estimates

For comparison with extant results, Table 2 presents full-sample (1962:01-2015:12) 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, the coefficients are significant across all maturities and forecasting models.

Bauer and Hamilton (2016) argue that prior findings of bond return predictability from non-yield factors based on conventional HAC standard errors are not robust due to the use of persistent predictor variables that are correlated with the innovations in bond returns. Instead, they find that the standard errors proposed by Ibragimov and Muller (2010) have excellent size and power properties in regressions where standard HAC inference is seriously distorted. Working with 12-month overlapping returns, we confirm Bauer and Hamilton's result and find little evidence of predictability from non-yield factors when based on the Ibragimov-Muller method. However, using one-month non-overlapping bond returns, we arrive at a very different conclusion as the evidence based on the Ibragimov-Muller p-values suggest that three of the eight Ludvigsson-Ng factors are statistically significant. These results suggest that the inference problems pointed out by Bauer and Hamilton (2016) largely disappear when using one-month non-overlapping bond returns rather than 12-month overlapping returns.[14]

Table 2 shows R^2 values in the range 1.6-2.1% for the model that uses FB as a predictor, 2.1-2.3% for the model that uses the CP factor, and around 4.6-5.2% for the model based on the LN factor. These values, which increase to 7-8% for the multivariate model, are notably smaller than those conventionally reported for the overlapping 12-month horizon. For comparison, at the one-year horizon we obtain R^2 values of 9-12%, 12-19%, and 13-17% for the FB, CP, and LN models, respectively.[15]

The extent of time variation in the parameter estimates of the multivariate FB-CP-LN model is displayed in Figure 2. All coefficients are notably volatile around 1980 and the coefficients continue to fluctuate throughout the sample.

To get a sense of the importance of parameter estimation error, Figure 3 plots full-sample posterior densities of the regression coefficients for the multivariate model that uses FB, CP and LN 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 precise estimates than the constant volatility benchmark, reflecting the ability of the SV model to reduce the weight on observations in highly volatile periods.

---

[14]Wei and Wright (2013) also 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. Their reverse regressions show that, even after accounting for such biases, bond excess returns still appear to be predictable.

[15]These values are a bit lower than those reported in the literature but are consistent with the range of results reported by Duffee (2013) . The weaker evidence reflects our use of an extended sample along with a tendency for the regression coefficients to decline towards zero at the end of the sample.

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 TVP-SV 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 have very 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, 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.

4.2   Calculation of out-of-sample Forecasts

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.[16] This experiment only uses information available at time 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 set to end at time t and re-estimate each period the principal components and the regression coefficients in equations (8) and (11).

We use 1962:01-1989:12 as our initial warm-up estimation sample and 1990:01-2015:12 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 (16) 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 $\mathcal{D}^t = \{rx_{\tau+1}^{(n)}, \boldsymbol{x}_\tau^{(n)}\}_{\tau=1}^{t-1} \cup \boldsymbol{x}_t^{(n)}$ are denoted $\left\{rx_{t+1,i}^{(n),j}\right\}$, $j = 1, ..., J$.[17]

---

[16]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.

[17]We run the Gibbs sampling algorithms recursively for all time periods between 1990:01 and 2015:12. At each point in 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 Appendix A.

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

$$p\left(rx_{t+1}^{(n)}\Big|\mathcal{D}^t\right) = \int p\left(rx_{t+1}^{(n)}\Big|\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)p\left(\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2}\Big|\mathcal{D}^t\right)d\mu d\boldsymbol{\beta}d\sigma_\varepsilon^{-2}. \qquad (28)$$

Return draws for the most general TVP-SV model are obtained from the predictive density[18]

$$\begin{aligned}
p\left(rx_{t+1}^{(n)}\Big|\mathcal{D}^t\right) &= \int p\left(rx_{t+1}^{(n)}\Big|\boldsymbol{\theta}_{t+1},h_{t+1},\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2},\mathcal{D}^t\right) \\
&\quad \times p\left(\boldsymbol{\theta}_{t+1},h_{t+1}\Big|\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2},\mathcal{D}^t\right) \qquad (29) \\
&\quad \times p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2}\Big|\mathcal{D}^t\right)d\mu d\boldsymbol{\beta}d\boldsymbol{\theta}^{t+1}d\boldsymbol{\gamma_\theta}d\boldsymbol{Q}dh^{t+1}d\lambda_0 d\lambda_1 d\sigma_\xi^{-2}.
\end{aligned}$$

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 (29). All Bayesian models integrate out uncertainty about the parameters.

4.3   Forecasting Performance

Although our models generate a full predictive distribution for bond returns, insights can be gained also from conventional point forecasts. To obtain point forecasts we first compute the posterior mean from the densities in (28) and (29). 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 $e_{t,i}^{(n)} = rx_t^{(n)} - \overline{rx}_{t,i}^{(n)}$, $t = \underline{t},...,\overline{t}$, where $\underline{t} = 1990:01$ and $\overline{t} = 2015:12$ 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

$$R_{OoS,i}^{(n)2} = 1 - \frac{\sum_{\tau=\underline{t}}^{\overline{t}} e_{\tau,i}^{(n)2}}{\sum_{\tau=\underline{t}}^{\overline{t}} e_{\tau,EH}^{(n)2}}. \qquad (30)$$

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, although these models fare better for the longer bond maturities. Conversely, almost all models estimated using our Bayesian approach generate significantly more accurate forecasts at either the 10% or 1% significance levels, using the test for equal predictive accuracy suggested by Clark and West (2007).

---

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

Similar results are obtained for the SV, TVP, and TVP-SV models which generate R^2_{OoS} values of 4-5% for the models that include the LN predictor.

Comparing R^2_{OoS} values across predictors, CP delivers the weakest results although the TVP-SV specification shows some evidence of predictive power for this variable, suggesting that the coefficient on CP varies over time. Conversely, the FB and, in particular, the LN predictor, add considerable improvements in out-of-sample predictive performance. To test the statistical significance of these differences, in results available in a web appendix, we perform pairwise comparisons across models with different predictor variables. Across all bond maturities and model specifications, we find that the R^2_{OoS} values are significantly higher for models that include the LN predictor compared to models that use either FB or CP.

Similarly, ranking the different model specifications across bond maturities and predictor variables, we find that the TVP-SV models produce the best out-of-sample forecasts in half of all cases with the SV model a distant second best. 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 simpler models. Even in cases where the TVP-SV model is not the best specification, it still performs nearly as well as the best model. In contrast, there are instances where the other models are clearly inferior to the TVP-SV model.

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:

$$\Delta CumSSE^{(n)}_{t,i} = \sum_{\tau=\underline{t}}^{t} \left(e^{(n)}_{\tau,EH} \right)^2 - \sum_{\tau=\underline{t}}^{t} \left(e^{(n)}_{\tau,i} \right)^2. \qquad (31)$$

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 2008, while the CP model underperforms between 1993 and 2006. In contrast, except for a few isolated months in 2002, 2008 and 2009, the LN model consistently beats the EH benchmark up to 2009, at which point its performance flattens against the EH model. A similar performance is seen for the multivariate model.

The predictive accuracy measures in (30) and (31) ignore information on the full probability distribution of returns. To evaluate the accuracy of the density forecasts obtained in (28) and (29), we compute the predictive likelihood (score) which gives a broad measure of accuracy of

density forecasts, see 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 (28)–(29) 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.[19] The results show that the SV and TVPSV models perform significantly better than the EH benchmark across all predictors and maturities. More modest but, in most cases, still significant improvements over the EH benchmark are observed for the linear and TVP specifications.

Figure 7 shows the cumulative log score (LS) differentials between the EH model and the ith model, computed analogously to (31) as

$$\Delta CumLS_{t,i} = \sum_{\tau=\underline{t}}^{t} \left[LS_{\tau,i} - LS_{\tau} \right]. \tag{32}$$

The dominant performance of the density forecasts generated by the SV and TVP-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.

4.4   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.

---

[19]To test if the differences in forecast accuracy are significant, we follow Clark and Ravazzolo (2015) and apply the Diebold and Mariano (1995) t-test for equality of the average log-scores based on the statistic $\overline{LS}_i = \frac{1}{\overline{t}-\underline{t}+1} \sum_{\tau=\underline{t}}^{\overline{t}} (LS_{\tau,i} - LS_{\tau,EH})$. 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.

5 Economic Value of Return Forecasts

So far our analysis concentrated on statistical measures of predictive accuracy. We next turn our attention to whether the apparent gains in predictive accuracy translate into better investment performance.

5.1 Bond Holdings

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, $\tilde{y}_t = y_t^{(1/12)}$. Under power utility

$$U\left(\omega_t^{(n)}, rx_{t+1}^{(n)}\right) = \frac{\left[\left(1 - \omega_t^{(n)}\right)\exp\left(\tilde{y}_t\right) + \omega_t^{(n)}\exp\left(\tilde{y}_t + rx_{t+1}^{(n)}\right)\right]^{1-A}}{1 - A}, \quad A > 0, \tag{33}$$

where A captures the investor's risk aversion.

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

$$\omega_t^{(n)*} = \arg\max_{\omega_t^{(n)}} \int U\left(\omega_t^{(n)}, rx_{t+1}^{(n)}\right) p\left(rx_{t+1}^{(n)}\Big| \mathcal{D}^t\right) drx_{t+1}^{(n)}. \tag{34}$$

The integral in (34) 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 (28) and (29). For each of the candidate models, i, we approximate the solution to (34) by

$$\widehat{\omega}_{t,i}^{(n)} = \arg\max_{\omega_{t,i}^{(n)}} \frac{1}{J}\sum_{j=1}^{J}\left\{\frac{\left[\left(1 - \omega_{t,i}^{(n)}\right)\exp\left(\tilde{y}_t\right) + \omega_{t,i}^{(n)}\exp\left(\tilde{y}_t + rx_{t+1,i}^{(n),j}\right)\right]^{1-A}}{1 - A}\right\}. \tag{35}$$

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

To make our results directly comparable to earlier studies such as Thornton and Valente (2012), we assume a coefficient of risk aversion of $A = 5$ and constrain the weights on each bond maturity to $-1 \leq \omega_{i,t} \leq 2$ $(i = 1, ..., 4)$, thus ruling out extreme allocations. Moreover, we also report results under mean-variance utility.

5.2 Multivariate asset allocation

So far we estimated univariate models separately for each bond maturity. We next generalize this to a multivariate setting where investors jointly model bond excess returns across the four

maturities. To this end, consider the following VAR(1) model

$$\boldsymbol{y}_{t+1} = \boldsymbol{c} + A_1 \boldsymbol{y}_t + \boldsymbol{\varepsilon}_{t+1}, \quad t = 1, ..., T-1, \tag{36}$$

where $\boldsymbol{y}_{t+1} = \left(rx_{t+1}^{(2)}, rx_{t+1}^{(3)}, rx_{t+1}^{(4)}, rx_{t+1}^{(5)}, \boldsymbol{x}_{t+1}' \right)'$ is an $m \times 1$ vector, $\boldsymbol{c}$ is an $m \times 1$ vector of intercepts, A_1 is an $m \times m$ matrix of coefficients on the lagged dependent variables, and $\boldsymbol{\varepsilon}_{t+1} \sim N\left(\boldsymbol{0}, \Omega_{t+1}\right)$, where Ω_{t+1} is an $m \times m$ covariance matrix. Factoring the covariance matrix as $\Omega_{t+1} = \Gamma^{-1} \Sigma_{t+1}^{1/2} \left(\Sigma_{t+1}^{1/2} \Gamma^{-1} \right)'$, we can write $\varepsilon_{t+1} = \Gamma^{-1} \Sigma_{t+1}^{1/2} \mathbf{u}_{t+1}$, with $\mathbf{u}_{t+1} \sim N\left(\boldsymbol{0}, I_m\right)$. Letting $h_{i,t+1}$ denote the i-th element of Σ_{t+1}, we specify the following law of motion for the log variances:

$$\ln h_{i,t+1} = \ln h_{i,t} + e_{i,t+1}, \quad i = 1, ..., m, \tag{37}$$

where the vector of innovations, $\mathbf{e}_{t+1} \sim N\left(\boldsymbol{0}, \Phi\right)$ is independent across time with variance matrix Φ as in Primiceri (2005). This gives us a multivariate SV model. To keep the analysis simple, and in view of the findings that stochastic volatility has a first-order effect on the results, we do not consider multivariate models with time-varying parameters.

Assuming that bond returns are joint lognormally distributed, following Campbell and Viceira (1999) we can approximate excess returns on a bond portfolio, $r_{p,t+1}$, by

$$r_{p,t+1} = \widetilde{y}_t + \boldsymbol{\omega}_t' \boldsymbol{rx}_{t+1} + \frac{1}{2} \boldsymbol{\omega}_t' \boldsymbol{\sigma}_{t+1|t}^2 - \frac{1}{2} \boldsymbol{\omega}_t' \Omega_{t+1|t}^{[1:4,1:4]} \boldsymbol{\omega}_t, \tag{38}$$

where $\boldsymbol{\omega}_t$ is a vector of portfolio weights, $\widetilde{y}_t = y_t^{(1/12)}$ denotes the one-month riskless T-bill rate, $\Omega_{t+1|t}^{[1:4,1:4]}$ denotes the top-left 4×4 partition of $\Omega_{t+1|t}$, the covariance matrix of the joint predictive density $p\left(\boldsymbol{rx}_{t+1}|\mathcal{D}^t\right)$, and σ_{t+1}^2 is a 4×1 vector containing the first four diagonal elements of $\Omega_{t+1|t}$. The optimal weights on the four bonds are given by

$$\boldsymbol{\omega}_t = \frac{1}{A} \left(\Omega_{t+1|t}^{[1:4,1:4]} \right)^{-1} \left[E\left(\boldsymbol{rx}_{t+1}|\mathcal{D}^t \right) + \boldsymbol{\sigma}_{t+1|t}^2/2 \right], \tag{39}$$

where $E\left(\boldsymbol{rx}_{t+1}|\mathcal{D}^t \right)$ denotes the mean of $p\left(\boldsymbol{rx}_{t+1}|\mathcal{D}^t \right)$, the predictive density of the vector of bond excess returns.

5.3   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. First, consider the results with a single risky bond shown in the left-most columns under power utility (Panel A) and mean-variance utility (Panel B), respectively. The CER values generally increase with the bond maturity. In 11 of 16 cases, the highest CER values are found for the TVP-SV models. For example, for the three-variable TVP-SV model the CER value increases from 0.52% ($n = 2$) to 2.82% ($n = 5$). To test if the annualized CER values are statistically greater

than zero we use a Diebold-Mariano test.[20] With the notable exception of the two-year bond maturity, most of the CER values for the SV and TVP-SV models are significantly higher than those generated by the EH benchmark.

To disentangle the sources of the gains in predictive performance that we uncover, for each choice of predictor variable (FB, CP, LN and $FB+CP+LN$) and each bond maturity (2, 3, 4, and 5 years) we run formal pairwise tests across the different model specifications. The results (reported in a web appendix) show that accounting for stochastic volatility is important and the improvement of both the SV and TVP-SV model over the LIN model are statistically significant in many instances. Further confirming the importance of modeling volatility dynamics, the LIN and TVP models that do not account for stochastic volatility produce lower predictive likelihood values than an EH model with stochastic volatility (EH-SV). Interestingly, however, the SV and TVP-SV models that account for time-varying risk premia continue to produce significantly higher CER values than the EH-SV benchmark.

The pairwise tests further confirm that the inclusion of the LN macro factor as a predictor makes an important difference. For each model specification, the LN factors is the single best predictor, and we generally find significant improvements when moving from specifications based on CP or FB to a model that includes LN.

Figure 8 plots cumulative CER values, computed relative to the EH benchmark, for the three-factor model. These graphs parallel the cumulated sum of squared error difference plots in (31), the key difference being that they show the cumulated risk-adjusted gains from using a particular model instead of the EH model. Across all bond maturities the cumulative CER value at the end of the sample exceeds 50 percent for all models.

Turning to the multi-asset allocation problem under the linear or SV specifications–results for which are shown in the right-most column in Table 5–we find again that allowing for stochastic volatility leads to substantial improvements in CER values, on the order of 1.1-2.0% per annum. Once again, the CER values are substantially higher once the LN predictor is included and, for such models, the multi-asset results improve upon the case with a single risky asset.

It is worth pointing out two limitations to the analysis above. First, the bond returns analyzed here are not fully tradable in the sense that they rely on interpolated yields which do not correspond exactly with traded market prices. Interpolation techniques are necessary to use because only irregularly spaced maturities are available for many bonds. This means that the

---

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

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

and

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

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

21

simulated trading results reported here, as well as in other studies, should be interpreted with caution.

Second, the CER values reported in Table 5 ignore transaction costs. However, when we allow for transaction costs we continue to see sizeable gains over the EH benchmark. For example, assuming a one-way transaction cost of 10 basis points, the CER value for the strategy that uses the TVP-SV model to predict bond excess returns is reduced from 0.52% to 0.18% for the two-year bond and from 2.82% to 2.46% for the five-year bond.

With these limitations in mind, we conclude that there is economic evidence that returns on 2-5 year bonds can be exploited using predictor variables proposed in the literature. Moreover, the best performing models allow for time-varying mean and volatility dynamics.

5.4   Comparison with Other Studies

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. Instead, Thornton and Valente (2012) find that the Sharpe ratios of their bond portfolios decrease when accounting for such effects through rolling window estimation. In contrast, we find that incorporating time-varying parameters and stochastic volatility in many cases improves bond portfolio performance.

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 quarterly and annual returns data. Compared with the monthly results, the quarterly and annual R^2_{OoS} values decline somewhat. At the quarterly horizon the univariate specification including the LN factor and the trivariate specification including FB, CP and LN, continue to perform well across the four bond maturities. The LN factor also performs well at the annual horizon, particularly for the bonds with longer maturities ($n = 4, 5$). 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 well captured by the LN predictor and is missed when using longer return horizons, thus helping to explain the difference between our results and those of Thornton and Valente (2012) and Dewachter et al. (2014).

The setup of Sarno et al. (2016) 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. (2016) compute expected excess returns in the context of an affine term structure model and also do not consider stochastic volatility or time vary-

ing parameters; (ii) different predictor variables: Sarno et al. (2016) use latent state variables extracted from their term structure model to predict bond excess returns; and, (iii) different estimation methodologies: Sarno et al. (2016) do not follow the same Bayesian methodology that we use here and thus ignore parameter uncertainty.[21]

Joslin and Le (2014) find that no-arbitrage term structure models that incorporate stochastic volatility factors face difficulties in matching yield dynamics under both the physical and risk-neutral probability measures. Given the focus of our study, we only model stochastic volatility under the physical measure and do not impose no-arbitrage conditions. In view of the challenges to affine term structure models from jointly matching the conditional means and variances of bond yields (Dai and Singleton (2002)), this may also help explain the difference from the results in Sarno et al. (2016) which are based on an afffine term structure model.

Another study that is closely related to ours is Barillas (2015) who uncovers the economic importance of using unspanned macroeconomic factors in a dynamic portfolio selection exercise. A notable difference between our paper and Barillas (2015) is that the latter provides in-sample evidence while our results are conducted out-of-sample.[22]

Duffee (2013) expresses concerns related to data mining when interpreting results for macro predictors whose effects are not underpinned by theory. A particular concern is that the strong results for the LN factor are sample specific. The sample used by Ludvigson and Ng (2009) ends in 2003:12. One way to address this concern is by inspecting the performance of the three-factor model in the subsequent sample, i.e., from 2004:01 to 2015:12. Figure 6 , Figure 7 and Figure 8 show that the prediction models continue to generate more accurate forecasts, higher CER values and higher log-density scores than the EH benchmark after 2003. Hence, the predictive power of the LN factor is not limited to the original sample used to construct this variable.

6   Economic Drivers of Bond Return Predictability and Portfolio Performance

We next conduct a set of tests designed to shed light on the economic drivers of bond return predictability and portfolio performance. First, we explore whether bond return predictability varies across the economic cycle. Next, we test implications for variation in risk premia of

---

[21]Both Thornton and Valente (2012) and Sarno et al. (2016) report performance using the Θ measure of Ingersoll et al. (2007). For comparison, we also computed performance results using this measure; findings are reported in a web appendix. We find that the results for the Θ measure are very similar to those obtained using the CER values shown here. Without a proper joint test, it is difficult to conclusively evaluate the finding in Sarno et al. (2016) that 13 ouf of 25 out-of-sample Θ estimates are positive. However, we note that their Θ estimates are relatively small, the highest Θ value being 0.81% (see Panel A of their Table 4).

[22]The CER values displayed in Table 7 of Barillas (2015) are computed by equating the value functions of two investors with different information sets: for the first, the information set only includes bond prices, while for the second it also includes macroeconomic variables. Barillas (2015) evaluates the value functions by simulation, and conducts rebalancing at the daily frequency whereas we use monthly rebalancing.

asset pricing models featuring habit persistence or learning dynamics. Third, using an ICAPM setting, we conduct a set of formal asset pricing tests to study whether variation in risk premia is driven by time-varying covariances between bond returns and innovations in the factors driving the pricing kernel. Fourth, we consider how the results from the portfolio analysis in Section 5 are related to uncertainty about the economy and biases in agents' subjective beliefs. Finally, we discuss whether unspanned risk factors help explain the predictive power of the LN macro factor and our portfolio allocation results.

6.1 Cyclical Variation 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. Similarly, Sarno et al. (2016) find that there are larger gains from predictability of bond returns during times with high macro uncertainty. These findings are important since they 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 which equals one in recessions and zero in expansions. Table 6 uses full-sample parameter estimates, 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, 1990-2015.

Table 6 shows that the R^2 values are generally higher during recessions than in expansions. Moreover, this finding is robust across model specifications and predictor variables, the only 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 relation 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 (30). 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,

$$H_0: \qquad E[\underbrace{e_{EH,0}^2 - e_{i,0}^2}_{\Delta_0}] = E[\underbrace{e_{EH,1}^2 - e_{i,1}^2}_{\Delta_1}] \tag{40}$$

$$H_1: \qquad E[e_{EH,0}^2 - e_{i,0}^2] < E[e_{EH,1}^2 - e_{i,1}^2].$$

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 $\Delta_0 = \Delta_1$.[23]

Outcomes from this test are indicated by stars in the recession columns of Table 6. With the notable exception of the univariate FB model, we find that not only is the fit of the bond return prediction models better in recessions than in expansions, but this difference is highly statistically significant in most cases.[24]

We also compute results that split the sample into recessions and expansions using the unemployment gap recession indicator of Stock and Watson (2010).[25] This indicator is computable in real time and so is arguably more relevant than the NBER indicator which gets released with a considerable lag. Using this alternative measure of recessions we continue to find that return predictability tends to be stronger in recessions than in expansions.

Cujean and Hasler (2016) provide a theoretical explanation of these patterns. In their analysis, investors assess uncertainty using different models which leads them to interpret the same news differently depending on economic conditions. Predictability is concentrated in bad times because this is when disagreement among investors tends to spike. To test whether disagreement is higher during recessions, we proxy disagreement with the cross-sectional dispersion (75th minus the 25th percentile) in quarterly forecasts of the 3-month T-bill rate from the Survey of Professional Forecasters. Regressing this proxy on a constant and the NBER recession index, the estimated slope coefficient is positive (0.11) with a t-stat of 2.02 which, in the context of the Cujean and Hasler model, is consistent with higher return predictability during recessions.

6.2 Variation in Risk Premia

Asset pricing models featuring habit persistence such as Campbell and Cochrane (1999) suggest that risk premia move counter-cyclically and that the Sharpe ratio of the aggregate stock market should be higher during recessions due to a reduced surplus consumption ratio. Wachter (2006) derives implications for bond risk premia and the term structure of interest rates in a setting with habit persistence.

---

[23]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, to 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.

[24]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.

[25]This measure is based on the difference between the current unemployment rate and a three-year moving average of past unemployment rates.

Creal and Wu (2016) extend the consumption-based framework of these papers by allowing for time-variation in both prices and quantities of risk and show that this can introduce counter-cyclical dynamics in bond risk premia. Moreover, habit formation in the model of Creal and Wu (2016) depends not only on past consumption but also on past inflation and their calibrated results suggest that inflation risk is an important driver of bond risk premia.[26]

To the extent that our forecasts of bond excess returns reflect time-varying risk premia, following these papers we would expect (i) higher Sharpe ratios in recessions; (ii) a negative correlation between economic growth and forecasts of bond risk premia; and (iii) a positive correlation between inflation and consumption growth risk–key drivers of bond risk premia in Creal and Wu (2016)–and forecasts of bond risk premia.

To test the first implication, 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 based on the full sample to ensure enough observations in recessions. With exception of the univariate FB regressions, the Sharpe ratios are substantially higher during recessions than in expansions.

Turning to the second implication, Panel A of Table 8 reports contemporaneous correlations between forecasts of two-year bond excess returns and current real GDP growth. Except for the models that use FB as a predictor, the correlations are negative and highly statistically significant. Thus, lower economic growth appears to be associated with expectations of higher bond excess returns as predicted by consumption-based models. Correlations are particularly strong for the LN macro factor which is sensitive to the economic cycle and also is the predictor that generates the highest economic gains.

To test the final implication, we show correlations between forecasts of two-year bond excess returns and expected consumption growth risk (Panel B in Table 8) or expected inflation risk (Panel C). We use a model-free approach to measure uncertainty about consumption growth and inflation by means of the interquartile range of one-quarter-ahead forecasts of consumption and consumer prices, respectively, obtained from the Survey of Professional Forecasters (SPF).

With exception of the FB model, we find strongly positive and, in most cases, highly significant correlations between uncertainty about consumption growth and future inflation on the one hand, and expected bond risk premia on the other. Moreover, the correlations are strongest for expected inflation risk, consistent with the finding in Creal and Wu (2016) that time-varying inflation risk is an important driver of bond risk premia.

6.2.1 Asset pricing models with learning dynamics

Giacoletti et al. (2016) show that time-varying risk premia can also arise in dynamic term

---

[26]Wright (2011) and Abrahams et al. (2013) also emphasize the importance of inflation risk to bond return dynamics.

structure models in which agents learn about the parameters of the data generating process or about the value of an unobserved state variable.[27]

To explore the relation between this type of dynamics and our analysis, we compute the correlation between our forecasts and those from Giacoletti et al. (2016) based on a learning rule that updates beliefs using the history of bond yields and disagreement among investors.[28] The results, presented in Panel D in Table 8, show quite large and mostly significant correlations between the forecasts in Giacoletti et al. (2016) and our forecasts. These results suggest that dynamic learning effects could account for some of our findings of return predictability in the bond market.

6.3 Multivariate ICAPM test

Next, we explore the extent to which movements in bond risk premia arise from time variation in the covariance between bond returns and shocks to the factors that enter in the pricing kernel. To do so, we use the multivariate ICAPM setting of Bali (2008) and Bali and Engle (2010). Specifically, we estimate a system of seemingly unrelated regressions for the four bond maturities using three instruments $(\Delta Inflation, \Delta Default, \Delta Term)$ that have been widely used to capture variation in the state of the economy and, hence, are likely to affect the pricing kernel:

$$
\begin{aligned}
rx_{i,t+1} &= \alpha_i + \beta \times cov(rx_{i,t+1}, r_{mkt,t+1}) + \gamma \times cov(rx_{i,t+1}, \Delta Inflation_{t+1}) \\
&\quad + \delta \times cov(rx_{i,t+1}, \Delta Default_{t+1}) + \theta \times cov(rx_{i,t+1}, \Delta Term_{t+1}) + \epsilon_{it+1}.
\end{aligned} \tag{41}
$$

Here $rx_{i,t+1}$ denotes the excess return on bond i, r_{mkt} is the value-weighted excess return on NYSE, AMEX and NASDAQ stocks, $\Delta Inflation_{t+1}$ is the change in the inflation rate computed from the consumer price index, $\Delta Default_{t+1}$ is the change in the difference between yields on BAA and AAA-rated bond portfolios, and $\Delta Term_{t+1}$ is the change in the term spread computed as the difference between yields on Treasury bonds with long and short maturities. Conditional covariances are computed using the dynamic conditional correlation (DCC) specification of Engle (2002).

Panel A of Table 9 reports slope estimates and t-statistics from the model in (41).[29] The coefficient on the conditional covariance between bond and stock returns (β) is positive but insignificant. Conversely, the coefficients on the conditional covariance between bond returns and changes to either inflation or the default spread are negative and significant. This is as we would expect: Assets whose returns are high when the default premium is unexpectedly high act as a hedge against bad economic states and, thus, should earn a lower risk premium, consistent

---

[27]Giacoletti et al. (2016) provide evidence that accounting for real time learning and belief dispersion improves forecasts of bond risk premia relative to a simple expectations hypothesis model of the term structure.

[28]We thank the authors for making this data available to us.

[29]As in Bali and Engle (2010), the slope coefficients β, γ, δ and θ, are pooled across the equations, while the intercepts (α_i) are allowed to differ across equations.

with a negative estimate of δ. Similarly, assets that act as an inflation hedge can be expected to earn a lower risk premium, consistent with a negative value for γ.

Finally, for each bond maturity and each model specification we computed the correlation between the ICAPM-fitted risk premium estimates based on equation (41) and the forecasts from our bond excess return equations, which is a key driver of the portfolio results. The results, reported in Panel B of Table 9, show that the correlation estimates generally fall in the range 0.10-0.30 and are statistically significant.

Taken together, these findings suggest that time-variation in the conditional covariance between bond excess returns and innovations to a set of widely-used economic state variables can account, at least in part, for movements in bond risk premia.

6.4 Economic Drivers of Portfolio Performance

So far we have seen that the risk premia generated by our prediction models are closely related to the state of the macroeconomy and proxies for macroeconomic uncertainty in particular. We next consider how the results from the portfolio analysis in Section 5 are related to macroeconomic uncertainty and agents' subjective beliefs about interest rates.

6.4.1 Macroeconomic uncertainty and portfolio performance

To relate time variation in bond risk premia to our earlier portfolio analysis, we compute the time-series correlation between the realized utility obtained in the portfolio analysis and our measures of GDP growth and inflation uncertainty.

The results, presented in panels E and F of Table 8, show that the realized utility from the portfolio analysis is positively and significantly correlated with inflation uncertainty. The correlation between realized utility and GDP growth uncertainty, while always positive, is smaller in magnitude and less statistically significant. Hence, the bond portfolios perform better in economic utility terms during times where macroeconomic uncertainty is high. This also tends to be times with higher recession risk and indicates the importance of inflation risk as a driver of portfolio performance.

6.4.2 Subjective forecasts of interest rates and portfolio performance

Using survey data on interest rate forecasts, Piazzesi et al. (2015) find that subjective risk premia are less volatile and less cyclical than statistical risk premia. The reason for the discrepancy is that survey forecasts of interest rates are made as if both the level and the slope of the yield curve are more persistent than under common statistical models.

Piazzesi et al. (2015) derive the following equation to construct subjective bond risk premia from survey data on interest rate forecasts:

$$E_t\left[rx_{t,t+h}^{(n)}\right] = E_t^*\left[rx_{t,t+h}^{(n)}\right] + (n-h)\left(E_t^*\left[i_{t+h}^{(n-h)}\right] - E_t\left[i_{t+h}^{(n-h)}\right]\right), \tag{42}$$

where $E_t\left[rx_{t,t+h}^{(n)}\right]$, the statistical premium, and $E_t\left[i_{t+h}^{(n-h)}\right]$, the statistical interest-rate expectation, are obtained from a VAR(1), and $E_t^*\left[i_{t+h}^{(n-h)}\right]$, the subjective interest-rate expectation, is obtained from the Blue Chip data.

To see whether the utility gains from our portfolio analysis might be related to biases in market participants' forecasts of future interest rates, we regress utility gains, computed relative to the EH benchmark, on the absolute difference between the subjective and the statistical interest rate forecasts, $|E_t^*[i_{t+h}^{(n-h)}] - E_t[i_{t+h}^{(n-h)}]|$.[30] Results from these regressions, reported in a web Appendix, show a mostly positive (and statistically significant at the 10% level or better) correlation between utility gains and differences in the subjective and statistical interest rate forecasts.

These findings suggest that the scope for turning bond return predictability into a portfolio strategy that enhances utility is larger during times with greater differences between statistical and subjective expectations of future interest rates. A possible interpretation of this finding is that biases in agents' beliefs about future yields are in part accountable for the possibility of increasing economic utility by exploiting the return predictability that we uncover.

6.5   Unspanned Macro Factors

Many studies use only information in the yield curve to predict bond excess returns so our finding that the LN macro factor improves such forecasts may seem puzzling. However, as discussed by Duffee (2013), a possible explanation is that the macro variables are hidden or unspanned risk factors which do not show up in the yield curve because their effect on expected future bond excess returns and expected future short rates work in opposite directions and so tend to cancel out in (5).[31] To see if this possibility holds up, we use Blue Chip survey forecasts of future short-term (one-year) yields to construct an estimate of the first term on the right hand side in (5), averaging, at each point in time, across the forecast horizons available from the Blue Chip survey.[32] To construct an estimate of the second term on the right hand side in (5), we use the forecasts from the models introduced in Section 3.

---

[30]We also tried using the squared difference, $\left(E_t^*[i_{t+h}^{(n-h)}] - E_t[i_{t+h}^{(n-h)}]\right)^2$ and found similar results.

[31]See Huang and Shi (2014) for a related analysis of unspanned macro risk factors.

[32]The Blue Chip survey forecasts are conducted for yields on US treasuries with maturities of 6 months, 1, 2, 5, 7, 10 and 30 years. The survey is run monthly and panel members provide forecasts of the average realization over a particular calendar quarter beginning with the current quarter and extending four to five quarters into the future. This implies that the forecast horizon depends on the month of the quarter in which the forecasts are formed. To equate forecast horizons throughout the sample, we use the interpolation method suggested by Chun (2011) and adopted in Giacoletti et al. (2016).

Consistent with the unspanned risk factor story, Panels A1 and B1 of Table 10 show that our forecasts of bond excess returns and survey expectations of future treasury yields are strongly negatively correlated with three-factor (FB-CP-LN) R^2 values ranging from 0.02 to 0.09 for $n = 2$ years and from 0.10 to 0.18 for $n = 5$ years.

To relate this finding to our portfolio exercise, we expand the analysis to regress mean-variance expected utilities on Blue Chip forecasts of future short yields. We show the results for the two-year and five-year bond maturities in Panels A2 and B2 of Table 10. We find a negative and highly statistically significant relation between expected future bond yields and expected utility, suggesting that periods where yields are expected to be low coincide with high risk premia and high expected future utility.

The final part of our analysis estimates bond risk premia by fitting an affine term structure model to the cross section of bond yields. In Appendix B we explain how we use the approach of Joslin et al. (2011) and Wright (2011) to fit term structure models with unspanned macro risks to compute bond risk premia. We use the resulting risk premium estimates to regress, for a given maturity n, the corresponding mean bond excess returns, $\overline{rx}_t$, on a constant and the corresponding risk premium estimates $\overline{rp}_t^{(n)}$,

$$\overline{rx}_t^{(n)} = \mu + \beta \overline{rp}_t^{(n)} + u_t, \tag{43}$$

where $\overline{rx}_t^{(n)}$ denotes the predicted bond excess-return and $\overline{rp}_t^{(n)}$ denotes the risk premium estimate. Table 11 reports the estimated coefficient β along with its t-statistics for the FB-CP-LN model. For all specifications, the estimated β coefficient has the right sign (positive) and it is statistically significant for the SV and TVP-SV models fitted to the two shortest bond maturities.

7 Model Combinations

In addition to parameter uncertainty, investors face model uncertainty along with the possibility that the best model may change over time, i.e., model instability. 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. Similar to diversification benefits obtained for asset portfolios, model combination tends to stabilize forecasts relative to forecasts generated by individual return prediction models.

Recent model combination approaches such as Bayesian model averaging and the optimal prediction pool of Geweke and Amisano (2011) allow the weights on individual forecasting models

to reflect their predictive accuracy. Such combination schemes can therefore accommodate time variations in the relative performance of different models. This matters if the importance of features such as time varying parameters and stochastic volatility dynamics changes over time.

A final 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), Elliott et al. (2013) and Pettenuzzo et al. (2014) 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 all possible models obtained by combining the FB, CP and LN predictors, estimated using the linear, SV, TVP and TVP-SV approaches.

7.1 Combination Schemes

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

$$p\left(rx_{t+1}^{(n)}\middle|\mathcal{D}^t\right) = \frac{1}{N}\sum_{i=1}^{N} p\left(rx_{t+1}^{(n)}\middle| M_i, \mathcal{D}^t\right), \tag{44}$$

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 (28) and (29). This approach does not allow the weights on different models to change over time as a result of differences in predictive accuracy.

We also consider Bayesian model averaging (BMA) weights:

$$p\left(rx_{t+1}^{(n)}\middle|\mathcal{D}^t\right) = \sum_{i=1}^{N} \Pr\left(M_i|\mathcal{D}^t\right) p\left(rx_{t+1}^{(n)}\middle| M_i, \mathcal{D}^t\right). \tag{45}$$

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

$$\Pr\left(M_i|\mathcal{D}^t\right) = \frac{\Pr\left(\mathcal{D}^t\middle| M_i\right)\Pr\left(M_i\right)}{\sum_{j=1}^{N}\Pr\left(\mathcal{D}^t\middle| M_j\right)\Pr\left(M_j\right)}. \tag{46}$$

$\Pr\left(\mathcal{D}^t\middle| M_i\right)$ 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 $\Pr\left(M_i\right) = 1/N$.[33]

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

---

[33]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:

$$p\left(rx_{t+1}^{(n)}\middle| \mathcal{D}^t \right) = \sum_{i=1}^{N} w_{t,i}^* \times p\left(rx_{t+1}^{(n)}\middle| M_i, \mathcal{D}^t \right). \tag{47}$$

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

$$\boldsymbol{w}_t^* = \arg\max_{\boldsymbol{w}_t} \sum_{\tau=1}^{t-1} \log\left[\sum_{i=1}^{N} w_{ti} \times S_{\tau+1,i} \right], \tag{48}$$

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 $\boldsymbol{w}_t^* \in [0,1]^N$. As $t \to \infty$ the weights in (48) 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 (45) and (48), these combination methods accommodate changes in the relative performance of the different models. This is empirically important as we shall see.

7.2 Empirical Findings

Table 12 presents statistical and economic measures of out-of-sample forecasting performance for the three combination schemes. The combinations generate similar R_{OoS}^2 values which range between 3.1% and 5.5%. In all cases, the forecast combinations perform better than what one would expect from simply selecting a single model at random.

The predictive likelihood tests shown in Panel B of Table 12 strongly reject the null of equal predictive accuracy relative to the EH model. Finally, the CER values range from 0.4% for the shortest bond maturity ($n = 2$) to 2-3% for the longest maturity ($n = 5$) with the optimal weights and BMA weights generally being better than those based on the EW combination.

8 Conclusion

We analyze predictability of excess returns on US 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 estimation error and model uncertainty.

We find evidence of both statistically and economically significant predictability in bond excess returns. This contrasts with the findings of Thornton and Valente (2012) who conclude

that the statistical evidence on bond return predictability fails to translate into economic return predictability. We find that such differences can be attributed to the importance of accounting for the information in (unspanned) macro factors along with modeling stochastic volatility and time-varying parameters in monthly bond excess returns. Consistent with unspanned risk factor models, forecasts of bond excess returns that incorporate information on macro variables are strongly negatively correlated with survey forecasts of future short term yields.

Our finding of economically significant return predictability in the US Treasury bond market can be understood in terms of two broad themes. First, it is possible that our forecasting models incorporate a larger information set or use more sophisticated methods than those adopted by investors to form yield expectations. The mostly positive correlation between utility gains from using our bond return forecasts and differences between subjective (survey) and statistical interest rate forecasts is consistent with this story.[34] The importance to bond return predictability of the composite Ludvigson-Ng macro factor and of stochastic volatility dynamics also makes this explanation more plausible than if, say, we had found that simpler predictors and simpler forecasting models accounted for the economic gains from return predictability.

Second, it is possible that we used the wrong preferences to evaluate the economic significance of return predictability, implicitly leading to a misspecified model for the market equilibrium. Consistent with this story, our bond return forecasts are strongly positively correlated with inflation uncertainty and negatively correlated with economic growth, suggesting that time varying risk premia could be an important driver of the results. The performance of portfolios formed using our bond excess return forecasts, also tends to be higher during times when macroeconomic uncertainty is high and risk premia could be higher than assumed.

While, ultimately, short of observing investors' preferences and beliefs, it is difficult to quantify the relative importance of these two explanations to our findings, our diagnostic tests suggest that both factors are at play.

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 12*(4), 708–729.

---

[34]See also the analysis in Cieslak (2016) which links predictability in bond returns from unspanned risk factors to evidence that investors systematically overestimate future real rates and underestimate unemployment in the early stage of recessions. Indeed, we also link evidence on bond return predictability to the economic cycle, finding that return predictability is significantly higher during recessions.

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.

Bali, T. G. (2008). The intertemporal relation between expected returns and risk. *Journal of Financial Economics 87*, 101–131.

Bali, T. G. and R. F. Engle (2010). The intertemporal capital asset pricing model with dynamic conditional correlations. *Journal of Monetary Economics 57*, 377–390.

Barillas, F. (2015). Can we exploit predictability in bond markets? *Working Paper*.

Bauer, D. M. and J. D. Hamilton (2016). Robust bond risk premia. *Working Paper*.

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.

Campbell, J. Y. and L. M. Viceira (1999). Consumption and portfolio decisions when expected returns are time varying. *The Quarterly Journal of Economics 114*(2), 433–495.

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

Chun, A. L. (2011). Expectations, bond yields, and monetary policy. *Review of Financial Studies 24*(1), 208–247.

Cieslak, A. (2016). Short-rate expectations and unexpected returns in treasury bonds. *Working Paper*.

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 (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics 30*(4), 551–575.

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., 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.

Creal, D. D. and J. C. Wu (2016). Bond risk premia in consumption-based models. *Chicago Booth Working Paper*.

Cujean, J. and M. Hasler (2016). Why does return predictability concentrate in bad times? *Working Paper*.

Dai, Q. and K. J. Singleton (2002). Expectations puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics 63*, 415–441.

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

Del Negro, M. and G. E. Primiceri (2015). Time varying structural vector autoregressions and monetary policy: A corrigendum. *The Review of Economic Studies*, forthcoming.

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. (2010). Sharpe ratios in term structure models. Wokring paper, John Hopkins University.

Duffee, G. (2013). Forecasting interest rates. In G. Elliot and A. Timmermann (Eds.), *Handbook of Economic Forecasting, volume 2*, pp. 385 – 426. North Holland.

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.

Elliott, G., A. Gargano, and A. Timmermann (2013). Complete subset regressions. *Journal of Econometrics 177*(2), 357–373.

Engle, R. F. (2002). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(339-350).

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. 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.

Giacoletti, M., K. Laursen, and K. J. Singleton (2016). Learning, dispersion of beliefs, and risk premiums in an arbitrage-free term structure model. *Working Paper*.

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.

Huang, J.-s. and S. Shi (2014). Determinants of bond risk premia. *Working Paper*.

Ibragimov, R. and U. K. Muller (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Stastistics 28*(4).

Ingersoll, J., M. Spiegel, W. Goetzmann, and I. Welch (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies 20*(5), 1503–1546.

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

Joslin, S. and A. Le (2014). Interest rate volatility and no-arbitrage affine term structure models. *Working Paper*.

Joslin, S., M. Priebsch, and K. J. Singleton (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance 69*(3), 1197–1233.

Joslin, S., K. J. Singleton, and H. Zhu (2011). A new perspective on gaussian dynamic term structure models. *Review of Financial Studies 24*(3), 926–970.

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.

Lin, H., C. Wu, and G. Zhou (2016). Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. *Working Paper*.

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

McCracken, M. and S. Ng (2015). FRED-MD: A monthly database for macroeconomic research. St. Louis Fed Working Paper 2015-012A.

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

Pettenuzzo, D., A. Timmermann, and R. Valkanov (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics 114*(3), 517–553.

Piazzesi, M., J. Salomao, and M. Schneider (2015, March). Trend and cycle in bond premia. *Working Paper*.

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 (2016). The economic value of predicting bond risk premia. *Journal of Empirical Finance 37*, 247 – 267.

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.

Stock, J. H. and M. W. Watson (2010). Modeling inflation after the crisis. Presented at the FRB Kansas city symposium on macroeconomic policy: post-crisis and risks ahead, Jackson Hole, Wyoming.

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.

Wachter, J. A. (2006). A consumption-based model of the term structure of interest rates. *Journal of Financial Economics 79*, 365–399.

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.

Appendix A   Bayesian estimation and predictions

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 (17), then turn to the SV model in (21)-(22), the TVP model in (23)-(24), and the general TVP-SV model in (25)-(27).

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 (18)-(20) with the likelihood function yields the following posteriors:

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

and

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

where

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

and

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

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 $(\mu, \boldsymbol{\beta}, \sigma_\varepsilon^{-2})$. Draws from the predictive density $p\left(rx_{t+1}^{(n)} \middle| \mathcal{D}^t\right)$ can then be obtained by noting that

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

A.2   Stochastic Volatility model[35]

The SV model requires specifying a joint prior for the sequence of log return volatilities, h^t, the parameters λ_0 and λ_1, and the error precision, σ_ξ^{-2}. Writing $p\left(h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}\right) =$

---

[35]See Pettenuzzo et al. (2014) for a description of a similar algorithm where the priors are modified to impose economic constraints on the model parameters.

$p\left(h^t\middle|\lambda_0,\lambda_1,\sigma_\xi^{-2}\right)p\left(\lambda_0,\lambda_1\right)p\left(\sigma_\xi^{-2}\right)$, it follows from (22) that

$$p\left(h^t\middle|\lambda_0,\lambda_1,\sigma_\xi^{-2}\right)=\prod_{\tau=1}^{t-1}p\left(h_{\tau+1}\middle|h_\tau,\lambda_0,\lambda_1,\sigma_\xi^{-2}\right)p\left(h_1\right),\tag{A-6}$$

with $h_{\tau+1}|h_\tau,\lambda_0,\lambda_1,\sigma_\xi^{-2}\sim\mathcal{N}\left(\lambda_0+\lambda_1h_\tau,\sigma_\xi^2\right)$. Thus, to complete the prior elicitation for $p\left(h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2}\right)$, we only need to specify priors for h_1, the initial log volatility, λ_0, λ_1, and σ_ξ^{-2}. We choose these from the normal-gamma family as follows:

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

$$\begin{bmatrix}\lambda_0\\\lambda_1\end{bmatrix}\sim\mathcal{N}\left(\begin{bmatrix}\underline{m}_{\lambda_0}\\\underline{m}_{\lambda_1}\end{bmatrix},\begin{bmatrix}\underline{V}_{\lambda_0}&0\\0&\underline{V}_{\lambda_1}\end{bmatrix}\right),\quad\lambda_1\in(-1,1),\tag{A-8}$$

and

$$\sigma_\xi^{-2}\sim\mathcal{G}\left(1/\underline{k}_\xi,\underline{v}_\xi\left(t-1\right)\right).\tag{A-9}$$

We set $\underline{k}_\xi=0.01$ and set the remaining hyperparameters in (A-7) and (A-9) at $\underline{k}_h=10$ and $\underline{v}_\xi=1$ to imply uninformative priors, thus allowing the data to determine the degree of time variation in the return volatility. Following Clark and Ravazzolo (2015) we set the hyperparameters to $\underline{m}_{\lambda_0}=0$, $\underline{m}_{\lambda_1}=0.9$, $\underline{V}_{\lambda_0}=0.25$, and $\underline{V}_{\lambda_0}=1.0e^{-4}$. This corresponds to setting the prior mean and standard deviation of the intercept to 0 and 0.5, respectively, and represents uninformative priors on the intercept of the log volatility specification and a prior mean of the AR(1) coefficient, λ_1, of 0.9 with a standard deviation of 0.01. This is a more informative prior that matches persistent dynamics in the log volatility process.

To obtain draws from the joint posterior distribution $p\left(\mu,\boldsymbol{\beta},h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2}\middle|\mathcal{D}^t\right)$ under the SV model, we use the Gibbs sampler to draw recursively from the following four conditional distributions:

1. $p\left(h^t\middle|\mu,\boldsymbol{\beta},\lambda_0,\lambda_1,\sigma_\xi^{-2},\mathcal{D}^t\right)$.

2. $p\left(\mu,\boldsymbol{\beta}\middle|h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2},\mathcal{D}^t\right)$.

3. $p\left(\lambda_0,\lambda_1\middle|\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right)$.

4. $p\left(\sigma_\xi^{-2}\middle|\mu,\boldsymbol{\beta},h^t,\lambda_0,\lambda_1,\mathcal{D}^t\right)$.

We simulate from each of these blocks as follows. Starting with $p\left(h^t\middle|\mu,\boldsymbol{\beta},\lambda_0,\lambda_1,\sigma_\xi^{-2},\mathcal{D}^t\right)$, we employ the algorithm of Kim et al. (1998).[36] Define $rx_{\tau+1}^{(n)*}=rx_{\tau+1}^{(n)}-\mu-\beta'\boldsymbol{x}_\tau^{(n)}$ and note that $r_{\tau+1}^{(n)*}$ is observable conditional on μ,β. Next, rewrite (21) as

$$rx_{\tau+1}^{(n)*}=\exp\left(h_{\tau+1}\right)u_{\tau+1}.\tag{A-10}$$

---

[36] We apply the correction to the ordering of steps detailed in Del Negro and Primiceri (2015).

Squaring and taking logs on both sides of (A-10) yields a new state space system that replaces (21)-(22) with

$$rx_{\tau+1}^{(n)**} = 2h_{\tau+1} + u_{\tau+1}^{**}, \tag{A-11}$$

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \tag{A-12}$$

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 to drawing from $p\left(h^t \middle| \mu, \boldsymbol{\beta}, \lambda_0, \lambda_1, \sigma_\xi^{-2}, s^t, \mathcal{D}^t\right)$ instead of $p\left(h^t \middle| \mu, \boldsymbol{\beta}, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right)$, where $s^t = \{s_2, ..., s_t\}$ 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-11)-(A-12) as a linear Gaussian state space model, making use of the following approximation,

$$u_{\tau+1}^{**} \approx \sum_{j=1}^{7} q_j \mathcal{N}\left(m_j - 1.2704, v_j^2\right), \tag{A-13}$$

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-13) implies

$$u_{\tau+1}^{**} \middle| s_{\tau+1} = j \sim \mathcal{N}\left(m_j - 1.2704, v_j^2\right), \tag{A-14}$$

where each state has probability

$$\Pr\left(s_{\tau+1} = j\right) = q_j. \tag{A-15}$$

Draws for the sequence of states s^t can be easily obtained, noting that each of its elements can be independently drawn from the discrete density defined by

$$\Pr\left(s_{\tau+1} = j \middle| \mu, \boldsymbol{\beta}, \lambda_0, \lambda_1, \sigma_\xi^{-2}, h^t, \mathcal{D}^t\right) = \frac{q_i f_{\mathcal{N}}\left(rx_{\tau+1}^{(n)**} \middle| 2h_{\tau+1} + m_j - 1.2704, v_j^2\right)}{\sum_{l=1}^{7} q_l f_{\mathcal{N}}\left(rx_{\tau+1}^{(n)**} \middle| 2h_{\tau+1} + m_l - 1.2704, v_l^2\right)}. \tag{A-16}$$

for $\tau = 1, ..., t-1$ and $j = 1, ..., 7$, and where $f_{\mathcal{N}}$ denotes the kernel of a normal density. Next, conditional on s^t, we can rewrite the nonlinear state space system as follows:

$$rx_{\tau+1}^{(n)**} = 2h_{\tau+1} + e_{\tau+1},$$

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \tag{A-17}$$

where $e_{\tau+1} \sim \mathcal{N}\left(m_j - 1.2704, v_j^2\right)$ with probability $\Pr\left(s_{\tau+1} = j \middle| \mu, \boldsymbol{\beta}, \lambda_0, \lambda_1, \sigma_\xi^{-2}, h^t, \mathcal{D}^t\right)$. 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.

41

Moving on to $p\left(\mu, \boldsymbol{\beta} \mid h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right)$, conditional on h^t it is straightforward to draw μ and $\boldsymbol{\beta}$ and apply standard results. Specifically,

$$\left[\begin{array}{c} \mu \\ \boldsymbol{\beta} \end{array}\right] \Bigg| h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t \sim \mathcal{N}\left(\overline{\boldsymbol{b}}, \overline{\boldsymbol{V}}\right), \tag{A-18}$$

with

$$\overline{\boldsymbol{V}} = \left\{\underline{\boldsymbol{V}}^{-1} + \sum_{\tau=1}^{t-1} \frac{1}{\exp\left(h_{\tau+1}\right)^2} \boldsymbol{x}_\tau^{(n)} \boldsymbol{x}_\tau^{(n)\prime}\right\}^{-1},$$

$$\overline{\boldsymbol{b}} = \overline{\boldsymbol{V}}\left\{\underline{\boldsymbol{V}}^{-1}\underline{\boldsymbol{b}} + \sum_{\tau=1}^{t-1} \frac{1}{\exp\left(h_{\tau+1}\right)^2} \boldsymbol{x}_\tau^{(n)} r x_{\tau+1}^{(n)}\right\}.$$

Next, the distribution $p\left(\lambda_0, \lambda_1 \mid \mu, \boldsymbol{\beta}, h^t, \sigma_\xi^{-2}, \mathcal{D}^t\right)$ takes the form

$$\lambda_0, \lambda_1 \mid \mu, \boldsymbol{\beta}, h^t, \sigma_\xi^{-2}, \mathcal{D}^t \sim \mathcal{N}\left(\left[\begin{array}{c} \overline{m}_{\lambda_0} \\ \overline{m}_{\lambda_1} \end{array}\right], \overline{\boldsymbol{V}}_\lambda\right) \times \lambda_1 \in (-1, 1),$$

where

$$\overline{\boldsymbol{V}}_\lambda = \left\{\left[\begin{array}{cc} \underline{V}_{\lambda_0}^{-1} & 0 \\ 0 & \underline{V}_{\lambda_1}^{-1} \end{array}\right] + \sigma_\xi^{-2} \sum_{\tau=1}^{t-1} \left[\begin{array}{c} 1 \\ h_\tau \end{array}\right] [1, h_\tau]\right\}^{-1}, \tag{A-19}$$

and

$$\left[\begin{array}{c} \overline{m}_{\lambda_0} \\ \overline{m}_{\lambda_1} \end{array}\right] = \overline{\boldsymbol{V}}_\lambda \left\{\left[\begin{array}{cc} \underline{V}_{\lambda_0}^{-1} & 0 \\ 0 & \underline{V}_{\lambda_1}^{-1} \end{array}\right]\left[\begin{array}{c} \underline{m}_{\lambda_0} \\ \underline{m}_{\lambda_1} \end{array}\right] + \sigma_\xi^{-2} \sum_{\tau=1}^{t-1} \left[\begin{array}{c} 1 \\ h_\tau \end{array}\right] h_{\tau+1}\right\}. \tag{A-20}$$

Finally, the posterior distribution for $p\left(\sigma_\xi^{-2} \mid \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \mathcal{D}^t\right)$ is readily available using

$$\sigma_\xi^{-2} \mid \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \mathcal{D}^t \sim \mathcal{G}\left(\left[\frac{\underline{k}_\xi \underline{v}_\xi (t-1) + \sum_{\tau=1}^{t-1}\left(h_{\tau+1} - \lambda_0 - \lambda_1 h_\tau\right)^2}{(1 + \underline{v}_\xi)(t-1)}\right]^{-1}, (1 + \underline{v}_\xi)(t-1)\right). \tag{A-21}$$

Draws from the predictive density $p\left(r x_{t+1}^{(n)} \mid \mathcal{D}^t\right)$ can be obtained by noting that

$$
\begin{aligned}
p\left(r x_{t+1}^{(n)} \mid \mathcal{D}^t\right) = & \int p\left(r x_{t+1}^{(n)} \mid h_{t+1}, \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right) \\
& \times p\left(h_{t+1} \mid \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right) \\
& \times p\left(\mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2} \mid \mathcal{D}^t\right) d\mu d\boldsymbol{\beta} dh^{t+1} d\lambda_0 d\lambda_1 d\sigma_\xi^{-2}.
\end{aligned}
\tag{A-22}
$$

The first term in the integral above, $p\left(r x_{t+1}^{(n)} \mid h_{t+1}, \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right)$, 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, $p\left(h_{t+1} \mid \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right)$, reflects how period $t+1$ volatility may drift away from h_t over

time. Finally, the last term in the integral, $p\left(\mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2} \middle| \mathcal{D}^t\right)$, measures parameter uncertainty in the sample.

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

1. Simulate from $p\left(\mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2} \middle| \mathcal{D}^t\right)$: draws from $p\left(\mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2} \middle| \mathcal{D}^t\right)$ are obtained from the Gibbs sampling algorithm described above.

2. Simulate from $p\left(h_{t+1} \middle| \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right)$: 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., $p\left(h_{t+1} \middle| \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right) = p\left(h_{t+1} \middle| h_t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right)$. Note also that (22) along with the distributional assumptions made on $\xi_{\tau+1}$ imply that

$$h_{t+1} \middle| h_t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t \sim \mathcal{N}\left(\lambda_0 + \lambda_1 h_t, \sigma_\xi^2\right). \tag{A-23}$$

3. Simulate from $p\left(rx_{t+1}^{(n)} \middle| h_{t+1}, \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right)$: For a given h_{t+1}, μ, and β, note that h^t, λ_0, λ_1, and σ_ξ^{-2} become redundant, i.e., $p\left(rx_{t+1}^{(n)} \middle| h_{t+1}, \mu, \boldsymbol{\beta}, h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}, \mathcal{D}^t\right) = p\left(rx_{t+1}^{(n)} \middle| h_{t+1}, \mu, \boldsymbol{\beta}, \mathcal{D}^t\right)$. Then use the fact that

$$rx_{t+1}^{(n)} \middle| h_{t+1}, \mu, \boldsymbol{\beta}, \mathcal{D}^t \sim \mathcal{N}\left(\mu + \boldsymbol{\beta}' \boldsymbol{x}_\tau^{(n)}, \exp\left(h_{t+1}\right)\right). \tag{A-24}$$

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 (23)-(24) requires eliciting a joint prior for the sequence of time varying parameters $\boldsymbol{\theta}^t = \{\boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_t\}$, the parameter vector $\boldsymbol{\gamma_\theta}$, and the variance covariance matrix $\boldsymbol{Q}$. For $[\mu, \boldsymbol{\beta}]'$ and σ_ε^2, we follow the same prior choices made for the linear model:

$$\begin{bmatrix} \mu \\ \boldsymbol{\beta} \end{bmatrix} \sim \mathcal{N}\left(\underline{\boldsymbol{b}}, \underline{\boldsymbol{V}}\right), \tag{A-25}$$

and

$$\sigma_\varepsilon^{-2} \sim \mathcal{G}\left(\left(s_{rx,t}^{(n)}\right)^{-2}, \underline{v}_0\left(t-1\right)\right). \tag{A-26}$$

Turning to $\boldsymbol{\theta}^t$, $\boldsymbol{\gamma_\theta}$, and $\boldsymbol{Q}$, we first write $p\left(\boldsymbol{\theta}^t, \boldsymbol{\gamma_\theta}, \boldsymbol{Q}\right) = p\left(\boldsymbol{\theta}^t \middle| \boldsymbol{\gamma_\theta}, \boldsymbol{Q}\right) p\left(\boldsymbol{\gamma_\theta}\right) p\left(\boldsymbol{Q}\right)$, and note that (24) along with the assumption that $\boldsymbol{\theta}_1 = \boldsymbol{0}$ implies

$$p\left(\boldsymbol{\theta}^t \middle| \boldsymbol{\gamma_\theta}, \boldsymbol{Q}\right) = \prod_{\tau=1}^{t-1} p\left(\boldsymbol{\theta}_{\tau+1} \middle| \boldsymbol{\theta}_\tau, \boldsymbol{\gamma_\theta}, \boldsymbol{Q}\right), \tag{A-27}$$

with $\boldsymbol{\theta}_{\tau+1} \middle| \boldsymbol{\theta}_\tau, \boldsymbol{\gamma_\theta}, \boldsymbol{Q} \sim \mathcal{N}\left(diag\left(\boldsymbol{\gamma_\theta}\right) \boldsymbol{\theta}_\tau, \boldsymbol{Q}\right)$. Thus, to complete the prior elicitation for $p\left(\boldsymbol{\theta}^t, \boldsymbol{\gamma_\theta}, \boldsymbol{Q}\right)$ we need to specify priors for $\boldsymbol{\gamma_\theta}$ and $\boldsymbol{Q}$.

43

We choose an Inverted Wishart distribution for $\mathbf{Q}$:

$$\mathbf{Q} \sim \mathcal{IW}\left(\underline{\mathbf{Q}}, \underline{v}_{\mathbf{Q}}\left(t_0 - 1\right)\right), \tag{A-28}$$

with

$$\underline{\mathbf{Q}} = \underline{k}_{\mathbf{Q}}\underline{v}_{\mathbf{Q}}\left(t_0 - 1\right)\underline{\mathbf{V}}. \tag{A-29}$$

$\underline{k}_{\mathbf{Q}}$ controls the degree of variation in the time-varying regression coefficients $\boldsymbol{\theta}_\tau$, with larger values of $\underline{k}_{\mathbf{Q}}$ implying greater variation in $\boldsymbol{\theta}_\tau$. Our analysis sets $\underline{k}_{\mathbf{Q}} = \left(\psi/100\right)^2$ and $\underline{v}_{\mathbf{Q}} = 10$. These are more informative priors than the earlier choices and limit the changes to the regression coefficients to be $\underline{\psi}/100$ on average.

We specify the elements of $\boldsymbol{\gamma_\theta}$ to be a priori independent of each other with generic element $\gamma_{\boldsymbol{\theta}}^i$

$$\gamma_{\boldsymbol{\theta}}^i \sim \mathcal{N}\left(\underline{m}_{\gamma_\theta}, \underline{V}_{\gamma_\theta}\right), \quad \gamma_{\boldsymbol{\theta}}^i \in (-1, 1), \quad i = 1, ..., k. \tag{A-30}$$

where $\underline{m}_{\gamma_\theta} = 0.8$, and $\underline{V}_{\gamma_\theta} = 1.0e^{-6}$, implying relatively high autocorrelations.

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

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

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

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

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

We simulate from each of these blocks as follows. Starting with θ^t, we focus on $p\left(\boldsymbol{\theta}^t \middle| \mu, \boldsymbol{\beta}, \sigma_\varepsilon^{-2}, \boldsymbol{\gamma_\theta}, \boldsymbol{Q}, \mathcal{D}^t\right)$. Define $rx_{\tau+1}^{(n)*} = rx_{\tau+1}^{(n)} - \mu - \beta'\boldsymbol{x}_\tau^{(n)}$ and rewrite (23) as follows:

$$rx_{\tau+1}^{(n)*} = \mu_\tau - \boldsymbol{\beta}_\tau'\boldsymbol{x}_\tau^{(n)} + \varepsilon_{\tau+1} \tag{A-31}$$

Knowledge of μ and $\boldsymbol{\beta}$ makes $rx_{\tau+1}^{(n)*}$ observable, and reduces (23) 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-31) using the algorithm of Carter and Kohn (1994).

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

$$\begin{bmatrix} \mu \\ \boldsymbol{\beta} \end{bmatrix} \middle| \sigma_\varepsilon^{-2}, \boldsymbol{\theta}^t, \boldsymbol{\gamma_\theta}, \boldsymbol{\gamma_\theta}, \boldsymbol{Q}, \mathcal{D}^t \sim \mathcal{N}\left(\overline{b}, \overline{V}\right), \tag{A-32}$$

and

$$\sigma_\varepsilon^{-2} \middle| \mu, \boldsymbol{\beta}, \boldsymbol{\theta}^t, \boldsymbol{Q}, \mathcal{D}^t \sim \mathcal{G}\left(\overline{s}^{-2}, \overline{v}\right), \tag{A-33}$$

44

where

$$\overline{V} = \left[\underline{V}^{-1} + \sigma_\varepsilon^{-2} \sum_{\tau=1}^{t-1} x_\tau^{(n)} x_\tau^{(n)\prime} \right]^{-1},$$

$$\overline{b} = \overline{V} \left[\underline{V}^{-1} \underline{b} + \sigma_\varepsilon^{-2} \sum_{\tau=1}^{t-1} x_\tau^{(n)} \left(rx_{\tau+1}^{(n)} - \mu_\tau - \beta_\tau' x_\tau^{(n)} \right) \right], \qquad \text{(A-34)}$$

$$\overline{s}^2 = \frac{\sum_{\tau=1}^{t-1} \left(rx_{\tau+1}^{(n)*} - \mu_\tau - \beta_\tau' x_\tau^{(n)} \right)^2 + \left(\left(s_{rx,t}^{(n)} \right)^2 \times \underline{v}_0 \, (t-1) \right)}{\overline{v}}, \qquad \text{(A-35)}$$

and $\overline{v} = (1 + \underline{v}_0)(t-1)$.

Next, obtaining draws from $p\left(\gamma_\theta \middle| \mu, \beta, \sigma_\varepsilon^{-2}, \theta^t, Q, \mathcal{D}^t \right)$ is straightforward. The $i-$th element γ_θ^i is drawn from the following distribution

$$\gamma_\theta^i \middle| \mu, \beta, \sigma_\varepsilon^{-2}, \theta^t, Q, \mathcal{D}^t \sim \mathcal{N}\left(\overline{m}_{\gamma_\theta}^i, \overline{V}_{\gamma_\theta}^i \right) \times \gamma_\theta^i \in (-1, 1) \qquad \text{(A-36)}$$

where

$$\overline{V}_{\gamma_\theta}^i = \left[\underline{V}_{\gamma_\theta}^{-1} + Q^{ii} \sum_{\tau=1}^{t-1} \left(\theta_\tau^i \right)^2 \right]^{-1},$$

$$\overline{m}_{\gamma_\theta}^i = \overline{V}_{\gamma_\theta}^i \left[\underline{V}_{\gamma_\theta}^{-1} \underline{m}_{\gamma_\theta} + Q^{ii} \sum_{\tau=1}^{t-1} \theta_\tau^i \theta_{\tau+1}^i \right], \qquad \text{(A-37)}$$

and Q^{ii} is the $i-$th diagonal element of Q^{-1}.

As for $p\left(Q \middle| \mu, \beta, \sigma_\varepsilon^{-2}, \theta^t, \gamma_\theta, \mathcal{D}^t \right)$, we have that

$$Q \middle| \mu, \beta, \sigma_\varepsilon^{-2}, \theta^t, \mathcal{D}^t \sim \mathcal{IW}\left(\overline{Q}, \overline{v}_Q \right), \qquad \text{(A-38)}$$

where

$$\overline{Q} = \underline{Q} + \sum_{\tau=1}^{t-1} \left(\theta_{\tau+1} - diag\left(\gamma_\theta \right) \theta_\tau \right) \left(\theta_{\tau+1} - diag\left(\gamma_\theta \right) \theta_\tau \right)'. \qquad \text{(A-39)}$$

and $\overline{v}_Q = \left(1 + \underline{v}_Q \right)(t-1)$.

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

$$p\left(rx_{t+1}^{(n)} \middle| \mathcal{D}^t \right) = \int p\left(rx_{t+1}^{(n)} \middle| \theta_{t+1}, \mu, \beta, \theta^t, \gamma_\theta, Q, \sigma_\varepsilon^{-2}, \mathcal{D}^t \right)$$

$$\times p\left(\theta_{t+1} \middle| \mu, \beta, \theta^t, \gamma_\theta, Q, \sigma_\varepsilon^{-2}, \mathcal{D}^t \right) \qquad \text{(A-40)}$$

$$\times p\left(\mu, \beta, \theta^t, \gamma_\theta, Q, \sigma_\varepsilon^{-2} \middle| \mathcal{D}^t \right) d\mu d\beta d\theta^{t+1} d\gamma_\theta dQ d\sigma_\varepsilon^{-2}.$$

The first term in the integral above, $p\left(rx_{t+1}^{(n)} \middle| \theta_{t+1}, \mu, \beta, \theta^t, \gamma_\theta, Q, \sigma_\varepsilon^{-2}, \mathcal{D}^t \right)$, 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, $p\left(\boldsymbol{\theta}_{t+1}|\,\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)$, reflects that the regression parameters may drift away from $\boldsymbol{\theta}_t$ over time. Finally, the last term in the integral, $p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},\sigma_\varepsilon^{-2}|\,\mathcal{D}^t\right)$, measures parameter uncertainty.

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

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

2. Simulate from $p\left(\boldsymbol{\theta}_{t+1}|\,\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)$: 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., $p\left(\boldsymbol{\theta}_{t+1}|\,\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},\sigma_\varepsilon^{-2},\mathcal{D}^t\right)=p\left(\boldsymbol{\theta}_{t+1}|\,\boldsymbol{\theta}_t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},\mathcal{D}^t\right)$. Note also that (24), along with the distributional assumptions made with regards to $\eta_{\tau+1}$, imply that

$$\boldsymbol{\theta}_{t+1}|\,\boldsymbol{\theta}_t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},\mathcal{D}^t \sim \mathcal{N}\left(diag\left(\boldsymbol{\gamma_\theta}\right)\boldsymbol{\theta}_t,\boldsymbol{Q}\right). \tag{A-41}$$

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

$$rx_{t+1}^{(n)}\Big|\,\boldsymbol{\theta}_{t+1},\mu,\boldsymbol{\beta},\sigma_\varepsilon^{-2},\mathcal{D}^t \sim \mathcal{N}\left(\left(\mu+\mu_t\right)+\left(\boldsymbol{\beta}+\boldsymbol{\beta}_t\right)'\boldsymbol{x}_\tau^{(n)},\sigma_\varepsilon^2\right). \tag{A-42}$$

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-25) and (A-26) for the regression parameters, (A-7) and (A-9) for the SV component, and (A-28) and (A-29) for the TVP component.

To obtain draws from the joint posterior distribution $p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2}\Big|\,\mathcal{D}^t\right)$ under the TVP-SV model, we use the Gibbs sampler to draw recursively from the following seven conditional distributions:

1. $p\left(\boldsymbol{\theta}^t|\,\mu,\boldsymbol{\beta},\lambda_0,\lambda_1,\sigma_\varepsilon^{-2},\boldsymbol{\gamma_\theta},\boldsymbol{Q},\mathcal{D}^t\right).$

2. $p\left(\mu,\boldsymbol{\beta},\lambda_0,\lambda_1,\sigma_\varepsilon^{-2}|\,\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},\mathcal{D}^t\right).$

3. $p\left(h^t|\,\mu,\boldsymbol{\beta},\lambda_0,\lambda_1,\sigma_\xi^{-2},\mathcal{D}^t\right).$

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

5. $p\left(\boldsymbol{Q}|\,\mu,\boldsymbol{\beta},\lambda_0,\lambda_1,\sigma_\varepsilon^{-2},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\mathcal{D}^t\right).$

6. $p\left(\lambda_0,\lambda_1|\,\mu,\boldsymbol{\beta},h^t,\sigma_\xi^{-2},\mathcal{D}^t\right).$

7. $p\left(\sigma_\xi^{-2}\middle|\mu,\boldsymbol{\beta},h^t,\lambda_0,\lambda_1,\mathcal{D}^t\right).$

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

$$
\begin{aligned}
p\left(rx_{t+1}^{(n)}\middle|\mathcal{D}^t\right) &= \int p\left(rx_{t+1}^{(n)}\middle|\boldsymbol{\theta}_{t+1},h_{t+1},\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2},\mathcal{D}^t\right)\\
&\quad \times p\left(\boldsymbol{\theta}_{t+1},h_{t+1}\middle|\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2},\mathcal{D}^t\right)\\
&\quad \times p\left(\mu,\boldsymbol{\beta},\boldsymbol{\theta}^t,\boldsymbol{\gamma_\theta},\boldsymbol{Q},h^t,\lambda_0,\lambda_1,\sigma_\xi^{-2}\middle|\mathcal{D}^t\right)d\mu d\boldsymbol{\beta} d\boldsymbol{\theta}^{t+1}d\boldsymbol{\gamma_\theta}d\boldsymbol{Q}dh^{t+1}d\lambda_0 d\lambda_1 d\sigma_\xi^{-2}.
\end{aligned}
\tag{A-43}
$$

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

Appendix B   Estimation of bond risk premia

In this appendix we describe how we estimate bond risk premia using a dynamic Gaussian affine term structure model. Yields are first collected in a vector, Y_t, which contains rates for J different maturities. The risk factors that determine the yields are denoted by Z_t; these include both macro factors (M_t) and yield factors ($F_t^{\mathcal{L}}$) extracted as the first $\mathcal{L}$ principal components of yields, i.e., $Z_t = \left(M_t', F_t^{\mathcal{L}\prime}\right)$. The macro factors can be unspanned, i.e., they are allowed to predict future yields without having additional explanatory power for current yields beyond the yield factors.

We begin by assuming that the continuously compounded nominal one-period interest rate r_t depends on the yield factors, but not on the macro factors,

$$
r_t = \delta_0 + \delta_{\mathcal{F}}' F_t^{\mathcal{L}} + 0_{\mathcal{M}}' M_t.
\tag{B-1}
$$

Next, the risk factors are assumed to follow a Gaussian vector autoregression (VAR) under the risk neutral probability measure:

$$
F_t^{\mathcal{L}} = \mu^{\mathbb{Q}} + \phi^{\mathbb{Q}} F_{t-1}^{\mathcal{L}} + 0_{\mathcal{L}\times\mathcal{M}}' M_{t-1} + \Sigma \varepsilon_t^{\mathbb{Q}},
\tag{B-2}
$$

where $\varepsilon_t^{\mathbb{Q}} \sim \mathcal{N}(0, I)$. Finally, the evolution in Z_t under the physical measure takes the form:

$$
Z_t = \mu + \phi Z_{t-1} + \Sigma \varepsilon_t,
\tag{B-3}
$$

where $\varepsilon_t \sim \mathcal{N}(0, I)$. Under these assumptions bond prices are exponentially affine in the yield factors and do not depend on the macro factors,

$$
P_t^{(n)} = \exp\left(\mathcal{A}_n + \mathcal{B}_n F_t^{\mathcal{L}} + 0_{\mathcal{M}\times\mathcal{M}}' M_t\right),
\tag{B-4}
$$

where $\mathcal{A} = \mathcal{A}\left(\mu^{\mathbb{Q}}, \phi^{\mathbb{Q}}, \delta_0, \delta_{\mathcal{F}}, \Sigma\right)$ and $\mathcal{B} = \mathcal{B}\left(\phi^{\mathbb{Q}}, \delta_{\mathcal{F}}\right)$ are affine loadings which are given by the following recursions:

$$\begin{aligned}
\mathcal{A}_{n+1} &= \mathcal{A}_n + \left(\mu^{\mathbb{Q}}\right)' \mathcal{B}_n + \frac{1}{2}\mathcal{B}_n'\Sigma\Sigma'\mathcal{B}_n - \delta_0 \\
\mathcal{B}_{n+1} &= \left(\phi^{\mathbb{Q}}\right)' \mathcal{B}_n - \delta_{\mathcal{F}}
\end{aligned} \tag{B-5}$$

with starting values $\mathcal{A}_0 = 0$ and $\mathcal{B}_0 = 0$.

Using these coefficients, the model-implied yields are obtained as

$$y_t^{(n)} = -\frac{1}{n}\log\left(P_t^{(n)}\right) = A_j + B_n F_t^{\mathcal{L}}, \tag{B-6}$$

where

$$A_n = -\frac{1}{n}\mathcal{A}_n, \qquad B_n = -\frac{1}{n}\mathcal{B}_n. \tag{B-7}$$

Similarly, the risk neutral price of a n-period bond, $\widetilde{P}_t^{(n)}$, and its implied yield, $\widetilde{y}_t^{(n)}$, can be calculated as

$$\widetilde{P}_t^{(n)} = \exp\left(\widetilde{\mathcal{A}}_n + \widetilde{\mathcal{B}}_n Z_t\right), \tag{B-8}$$

and

$$\widetilde{y}_t^{(n)} = \widetilde{A}_n + \widetilde{B}_n Z_t, \tag{B-9}$$

where

$$\widetilde{A}_n = -\frac{1}{n}\widetilde{\mathcal{A}}_n, \qquad \widetilde{B}_n = -\frac{1}{n}\widetilde{\mathcal{B}}_n. \tag{B-10}$$

and where $\widetilde{\mathcal{A}}_n$ and $\widetilde{\mathcal{B}}_n$ are given by the following recursions:

$$\begin{aligned}
\widetilde{\mathcal{A}}_{n+1} &= \widetilde{\mathcal{A}}_n + \mu'\widetilde{\mathcal{B}}_n + \frac{1}{2}\widetilde{\mathcal{B}}_n'\Sigma\Sigma'\widetilde{\mathcal{B}}_n - \delta_0 \\
\widetilde{\mathcal{B}}_{n+1} &= \phi'\widetilde{\mathcal{B}}_j - \delta_{\mathcal{F}}
\end{aligned} \tag{B-11}$$

initialized at $\widetilde{\mathcal{A}}_0 = 0$ and $\widetilde{\mathcal{B}}_0 = 0$.

We follow Joslin et al. (2011) and Wright (2011) and include in $F_t^{\mathcal{L}}$ the first three principal components of zero-coupon bond yields using maturities ranging from three months to ten years. As macro factors, M_t, we use exponentially weighted moving averages of monthly inflation and IPI growth. The data used for estimation are monthly yields on zero-coupon bonds, inflation, and IPI growth, from 1982:01 to 2015:12. Our estimation approach also follows closely Joslin et al. (2011). Using the resulting estimates of the model parameters, we compute the risk premium at all maturities as the difference between the yields computed under the risk neutral measure, $\mathbb{Q}$, and the yields calculated under the physical measure, i.e.,

$$rp_t^{(n)} = y_t^{(n)} - \widetilde{y}_t^{(n)}. \tag{B-12}$$

Figure 1. **Bond excess returns**
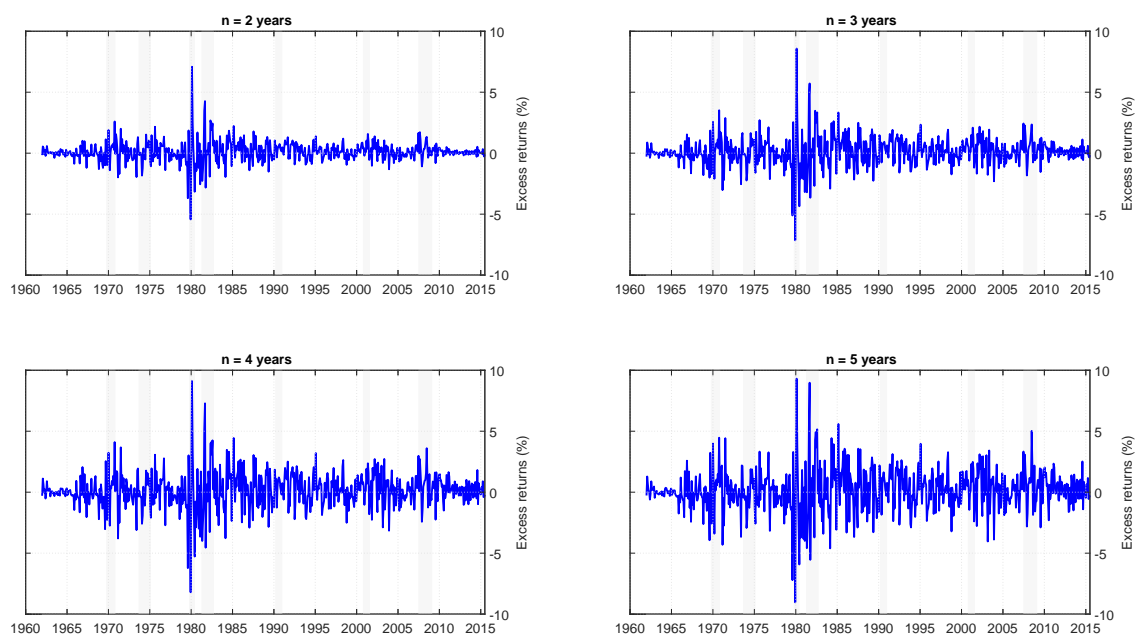
This figure shows time series of monthly bond excess returns (in percentage terms) for maturities (n) ranging from 2 years through 5 years. Monthly bond excess returns, $rx_{t+1/12}^{(n)}$, are computed from monthly yields, $y_t^{(n)}$, and are expressed in deviations from the 1-month T-bill rate, $rx_{t+1/12}^{(n)} = r_{t+1/12}^{(n)} - (1/12)y_t^{1/12}$, 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 2015.

Figure 2. **Parameter estimates for bond return forecasting model**
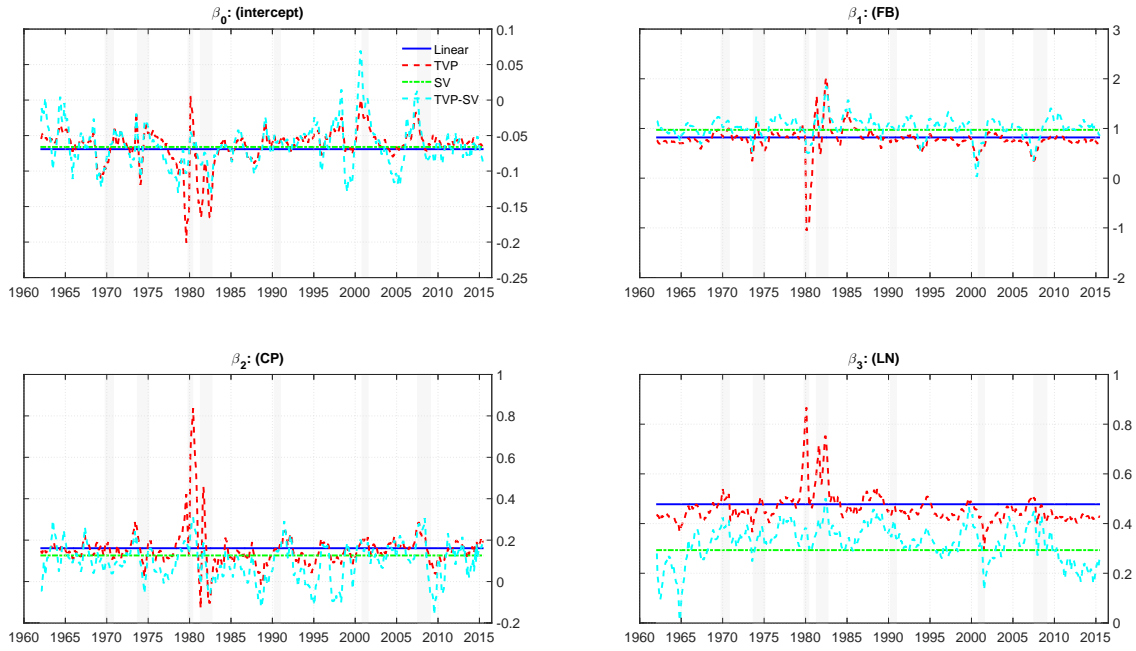
This figure displays parameter estimates for the FB-CP-LN model used to forecast monthly 3-year bond excess returns using as predictors the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) variables. The blue solid line represents the linear, constant coefficient model (Linear); the red 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 2015 and the parameter estimates are based on full-sample information.

Figure 3. **Posterior densities for model parameters**
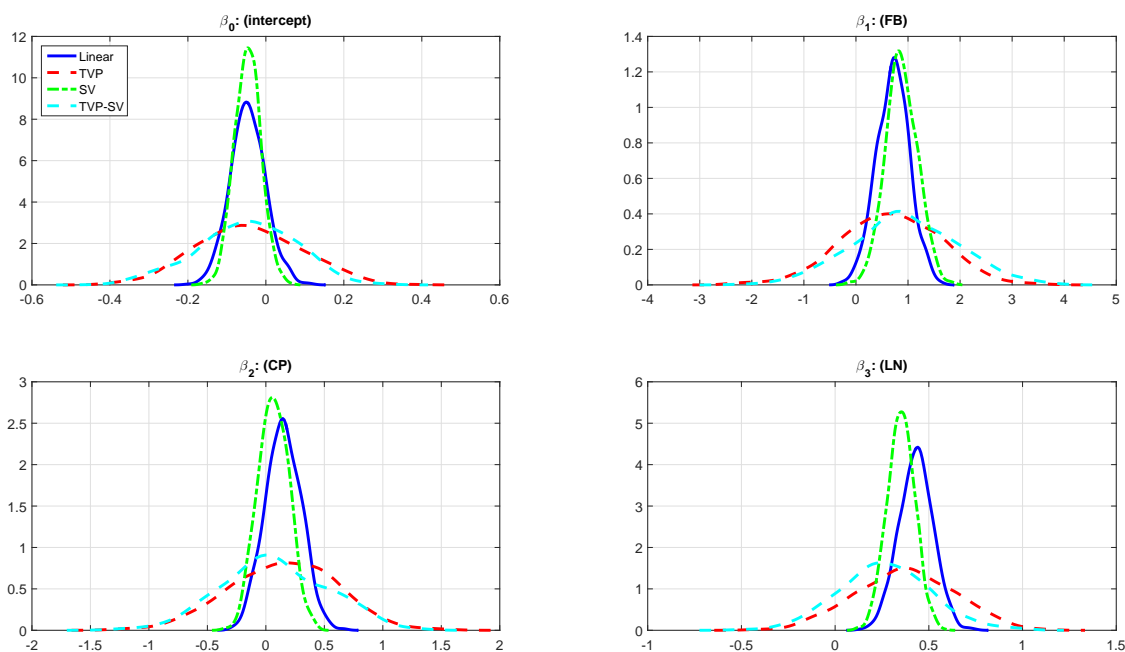
This figure displays posterior densities for the coefficients of the FB-CP-LN return model fitted to 3-year Treasury bonds, using as predictors the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) factors. The blue solid line represents the linear, constant coefficient (Linear) model; the 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 2015:12.

Figure 4. **Posterior densities for bond returns**
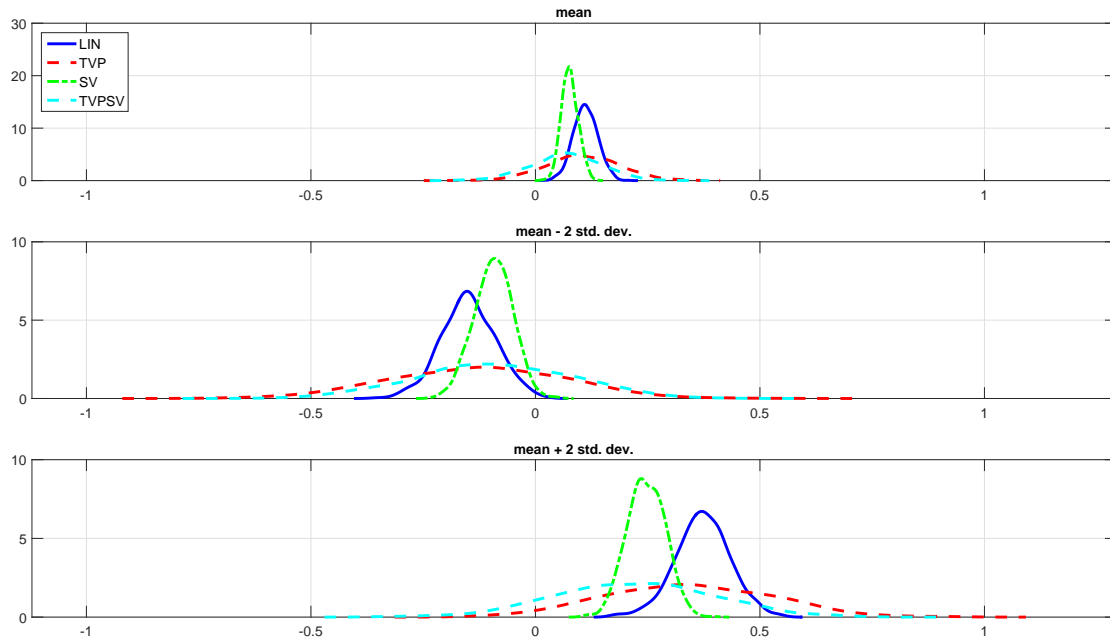
This figure shows the posterior density for excess returns on a three-year Treasury bond using the univariate Ludvigson-Ng (LN) state variable as a predictor. The LN variable is set at its sample mean $\overline{LN}$ (top panel), $\overline{LN} - 2stdev\,(LN)$ (middle panel), and $\overline{LN} + 2stdev\,(LN)$ (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 2015.

Figure 5. **Conditional mean and volatility estimates for bond excess returns**
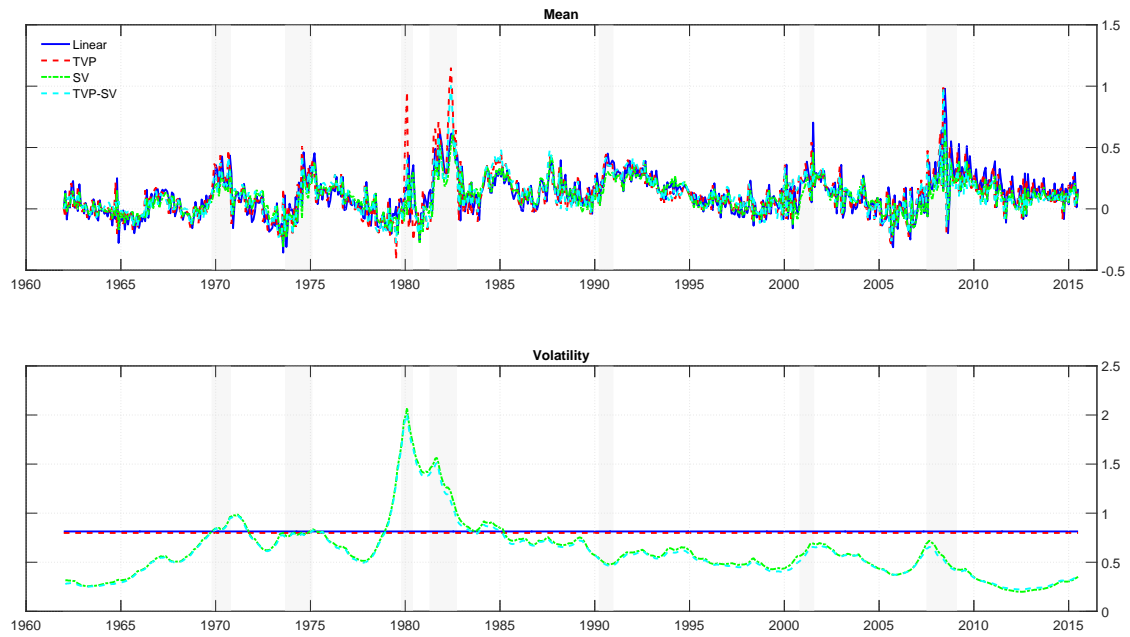
The top panel shows time-series of expected bond excess returns obtained from a range of models used to forecast monthly returns on a three-year Tresury bond using as predictors the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) factors. The blue solid line represents the linear, constant 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 2015 and the estimates are based on full-sample information.

Figure 6. **Cumulative sum of squared forecast error differentials**
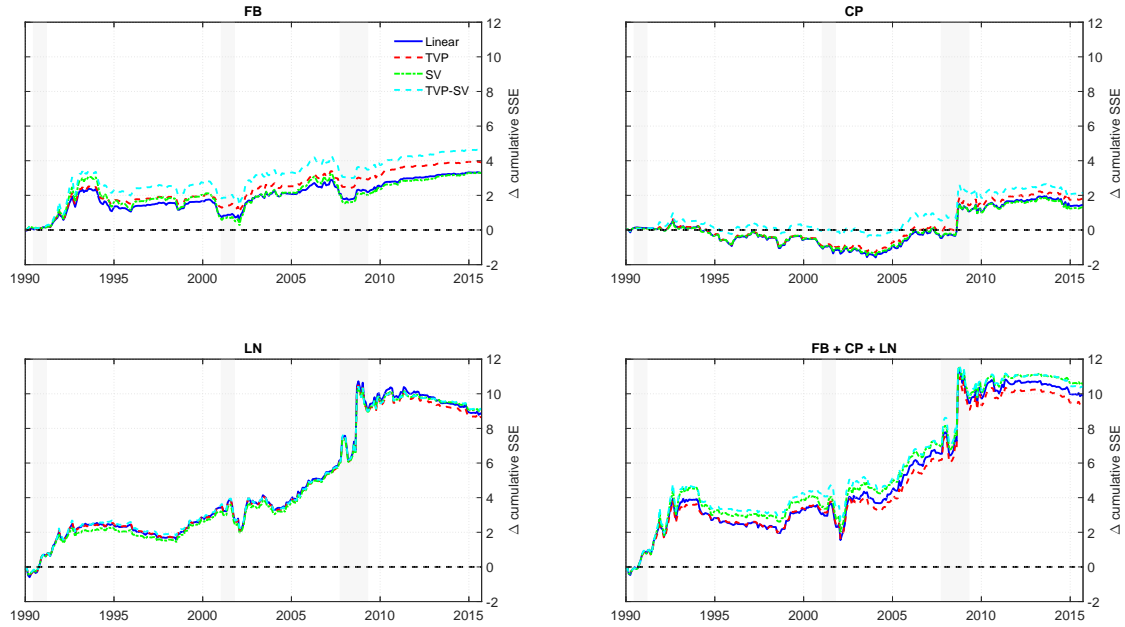
This figure shows the recursively calculated sum of squared forecast errors for the expectations hypothesis (EH) model minus the sum of squared forecast errors for a forecasting model with time-varying expected returns for a bond with a two year maturity, $(n = 2)$. Each month we recursively estimate 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 - 2015:12.

Figure 7. **Cumulative sum of log-score differentials**
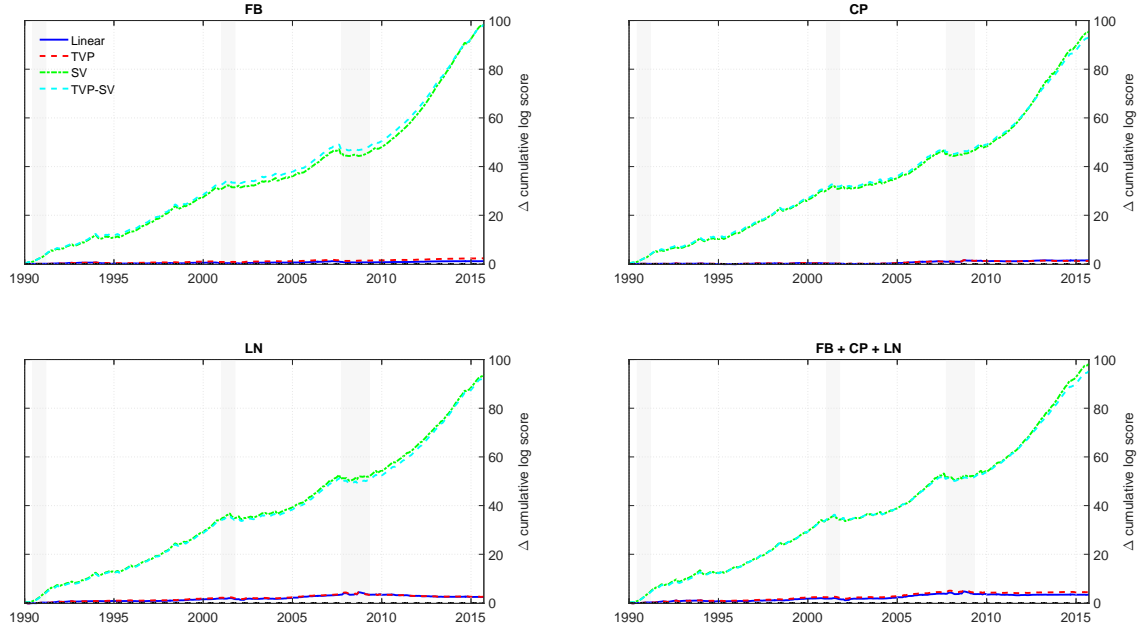
This figure shows the recursively calculated sum of log predictive scores from forecasting models with time-varying predictors minus the corresponding sum of log predictive scores for the EH model, using a 2-year Treasury bond. Each month we recursively estimate the parameters of the forecasting models 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 - 2015:12.

Figure 8. **Economic value of out-of-sample bond return forecasts**
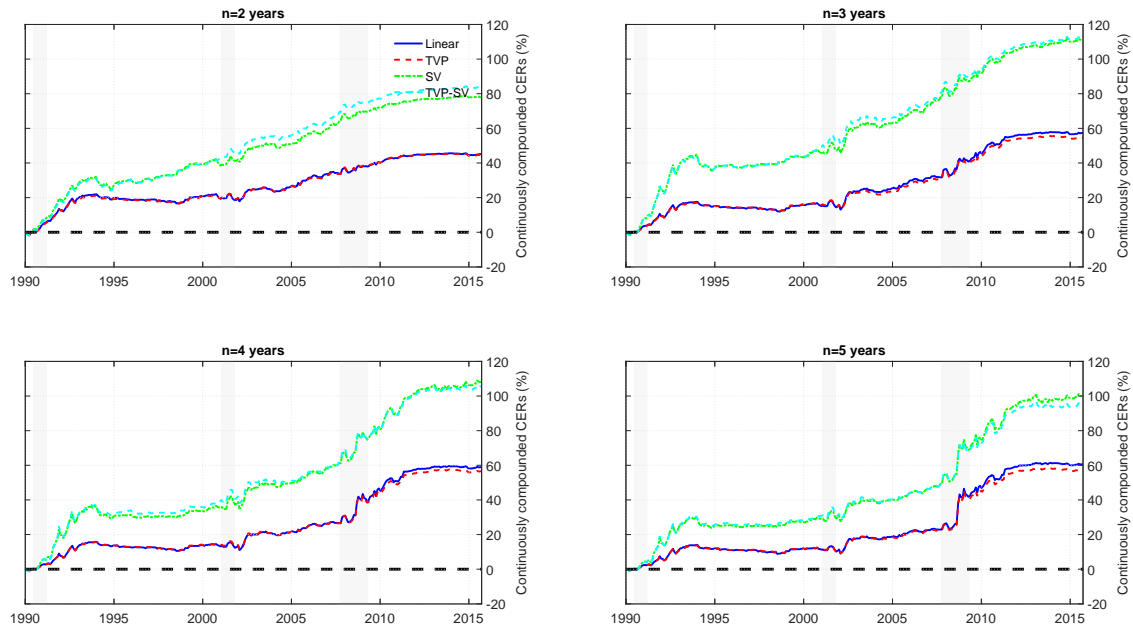
This figure plots cumulative certainty equivalent returns for the three-factor FB-CP-LN forecasting model that uses the Fama-Bliss (FB), Cochrane-Piazzesi (CP), and Ludvigson-Ng (LN) factors as predictors, measured relative to the expectations hypothesis (EH) model. Each month we compute the optimal 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 [−200%, 300%]. 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 - 2015:12.

Table 1. **Summary Statistics.**

	2 years	3 years	4 years	5 years
Panel A.1: One-month excess returns				
mean	1.3479	1.6679	1.9255	2.1353
mean (gross)	5.9985	6.3186	6.5761	6.7859
st.dev.	2.8636	4.0133	5.0528	6.0457
skew	0.5360	0.2268	0.0674	0.0223
kurt	15.9574	11.3365	8.3208	6.8606
AC(1)	0.1681	0.1496	0.1323	0.1162
Panel A.2: 12-month overlapping excess returns				
mean	0.4992	0.8602	1.1440	1.3757
mean (gross)	5.8838	6.2448	6.5285	6.7603
st.dev.	1.6954	3.0798	4.2830	5.3784
skew	-0.0772	-0.0628	-0.0337	-0.0053
kurt	4.0286	3.8186	3.6981	3.6572
AC(1)	0.9312	0.9319	0.9320	0.9312
Panel A.3: 12-month overlapping excess returns				
(Cochrane-Piazzesi)				
mean	0.4700	0.8670	1.1805	1.2927
mean (gross)	5.8511	6.2481	6.5616	6.6738
st.dev.	1.7178	3.1528	4.3924	5.4264
skew	0.0774	-0.0262	0.0004	-0.0115
kurt	3.6801	3.7473	3.6514	3.5716
AC(1)	0.9311	0.9333	0.9323	0.9236

Panel B: Predictors

	Fama Bliss				CP	LN
	2-years	3-years	4-years	5-years		
mean	0.1054	0.1287	0.1473	0.1623	0.1480	0.1480
st.dev.	0.0967	0.1120	0.1241	0.1339	0.1982	0.2887
skew	-0.0120	-0.2446	-0.2693	-0.2130	0.6316	0.6962
kurt	3.9157	3.5445	3.1892	2.8765	4.4039	4.8691
AC(1)	0.8801	0.8998	0.9130	0.9233	0.7073	0.3899

Panel C: Correlation Matrix

	FB-2	FB-3	FB-4	FB-5	CP	LN
FB-2	1.000	0.969	0.914	0.860	0.487	-0.121
FB-3		1.000	0.985	0.955	0.497	-0.073
FB-4			1.000	0.992	0.508	-0.026
FB-5				1.000	0.510	0.012
CP					1.000	0.126
LN						1.000

This table reports summary statistics for monthly bond excess returns and the predictor variables used in our study. Panels A.1-A.3 report the mean, standard deviation, skewness, kurtosis and first-order autocorrelation (AC(1)) of bond excess returns for 2 to 5-year bond maturities. Panel A.1 is based on monthly returns computed in excess of a one-month T-bill rate while Panels A.2 and Panel A.3 are based on 12-month overlapping returns, computed in excess of a 12-month T-bill rate. Gross returns do not subtract the risk-free rate. In Panels A.1 and A.2 returns are constructed using daily treasury yield data from Gurkaynak et al. (2007) while in Panel A.3 returns are constructed as in Cochrane and Piazzesi (2005) using the Fama-Bliss CRSP files. 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 1962-2015.

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

	FB	CP	LN	FB+CP+LN
		2 years		
β_{FB}	1.1724***			1.1823***
β_{CP}		0.6477**		0.2403
β_{LN}			0.6640***	0.6912***
R^2	0.0173	0.0226	0.0522	0.0795
		3 years		
β_{FB}	1.3803**			1.2338**
β_{CP}		0.8741**		0.3615
β_{LN}			0.9050***	0.9089***
R^2	0.0163	0.0208	0.0493	0.0718
		4 years		
β_{FB}	1.6639**			1.3368**
β_{CP}		1.1079**		0.4835
β_{LN}			1.1180***	1.0910***
R^2	0.0185	0.0211	0.0474	0.0694
		5 years		
β_{FB}	1.9555**			1.4330**
β_{CP}		1.3702**		0.6479
β_{LN}			1.3130***	1.2489***
R^2	0.0210	0.0227	0.0456	0.0684

 This table reports OLS estimates of the slope coefficients for four 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, column (4) 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 p-values computed using the Ibragimov and Muller (2010) procedure with q, the number of sample partitions, equal to 16. ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. The sample period is 1962-2015.

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

| | Panel A: 2 years | | | | |
Model	OLS	LIN	SV	TVP	TVPSV
FB	0.45%**	1.02%*	0.65%*	1.86%***	**2.82%***
CP	-2.27%	0.35%*	0.59%*	0.74%***	**1.76%**
LN	-4.30%	3.43%***	5.11%***	3.89%***	**5.75%***
FB+CP+LN	-5.64%	3.78%***	5.39%***	4.59%***	**6.24%***

| | Panel B: 3 years | | | | |
Model	OLS	LIN	SV	TVP	TVPSV
FB	2.02%**	1.62%**	1.61%**	1.92%***	**2.27%***
CP	-0.95%	0.71%*	0.65%*	0.89%*	**1.06%**
LN	0.49%***	4.34%***	4.46%***	4.24%***	**4.48%***
FB+CP+LN	-1.21%	4.86%***	**5.19%***	4.62%***	5.10%***

| | Panel C: 4 years | | | | |
Model	OLS	LIN	SV	TVP	TVPSV
FB	**2.28%***	1.58%***	1.65%**	1.71%***	1.82%**
CP	-0.53%	0.55%	0.43%	0.54%	**0.69%**
LN	2.15%***	**3.75%***	3.63%***	3.55%***	3.54%***
FB+CP+LN	0.36%***	4.51%***	**4.61%***	4.29%***	4.40%***

| | Panel D: 5 years | | | | |
Model	OLS	LIN	SV	TVP	TVPSV
FB	**2.50%***	1.61%***	1.67%**	1.63%**	1.72%**
CP	0.02%	0.70%*	0.59%	**0.72%**	0.64%
LN	3.17%***	**3.41%**	3.27%**	3.38%***	3.15%***
FB+CP+LN	1.30%***	4.22%***	**4.29%***	3.99%***	4.04%***

This table reports out-of-sample R^2 values for four 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: $R^2_{OoS} = 1 - \frac{\sum_{\tau=t-1}^{\bar{t}-1}(rx_{t+1} - \hat{r}_{xt+1|t})^2}{\sum_{\tau=t-1}^{\bar{t}-1}(rx_{t+1} - \bar{r}_{xt+1|t})^2}$ 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: $rx_{t+1} = \mu + \beta'x_t + \varepsilon_{t+1}$. $\bar{r}_{t+1|t}$ 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 2015. 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. For every model and maturity, we denote in bold font the R^2_{OoS} of the estimation method (LIN, SV, TVP and TVPSV) which delivers the best result.

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	0.004	**0.315*****	0.008**	0.315***	0.004**	**0.187*****	0.007**	0.186***
CP	0.005*	**0.307*****	0.004*	0.300***	0.005**	**0.184*****	0.006*	0.182***
LN	0.008*	**0.301*****	0.008**	0.298***	0.010**	**0.188*****	0.010**	0.186***
$FB + CP + LN$	0.011*	**0.315*****	0.014*	0.306***	0.013**	**0.193*****	0.014**	0.189***

| | Panel C: 4 years | | | | Panel D: 5 years | | | |
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.004**	**0.124*****	0.005**	0.123***	0.006**	0.091***	0.006**	**0.091*****
CP	0.004*	0.122***	0.004*	**0.123*****	0.004**	**0.089*****	0.006*	0.089***
LN	0.011**	**0.129*****	0.010**	0.128***	0.011**	**0.096*****	0.011*	0.095***
$FB + CP + LN$	0.014**	**0.134*****	0.015**	0.131***	0.015**	**0.101*****	0.015**	0.099***

This table reports the log predictive score for four forecasting models that allow for time-varying predictors relative to the log-predictive score computed under the expectation hypothesis (EH) model. The four 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 the combination 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 - 2015. ***: significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. For every model and maturity, we denote in bold font the Predictive Likelihood of the estimation method (LIN, SV, TVP and TVPSV) which delivers the best result.

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

Panel A: Power Utility									
	Univariate								Joint
	Panel A.1: 2 years				Panel A.2: 3 years				
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV	LIN
FB	-0.45%	-0.63%	-0.23%	**-0.05%**	0.16%	0.21%	0.29%	**0.59%***	-0.883
CP	-0.40%	-0.07%	-0.30%	**0.02%**	-0.32%	0.47%*	-0.23%	**0.66%****	-1.097
LN	**0.25%**	0.18%	0.25%	0.25%	1.14%***	**1.37%*****	1.11%***	1.28%***	2.298***
$FB+CP+LN$	0.22%	0.32%	0.29%	**0.52%****	1.12%***	1.53%***	1.18%***	**1.76%*****	1.501***
	Panel A.3: 4 years				Panel A.4: 5 years				
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV	SV
FB	0.76%*	1.02%*	0.88%**	**1.26%****	1.15%**	1.69%**	1.14%**	**1.73%****	0.481*
CP	-0.13%	0.62%	-0.06%	**0.87%****	0.19%	**0.85%****	0.20%	0.82%*	0.581*
LN	1.48%***	2.10%***	1.47%***	**2.14%*****	1.54%***	2.46%***	1.57%***	**2.49%*****	3.331***
$FB+CP+LN$	1.80%***	**2.34%*****	1.78%***	2.34%***	1.96%***	**2.92%*****	1.96%***	2.82%***	3.489***
Panel B: Mean Variance Utility									
	Univariate								Joint
	Panel B.1: 2 years				Panel B.2: 3 years				
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV	LIN
FB	-0.43%	-0.57%	-0.21%	**-0.01%**	0.22%	0.34%	0.35%	**0.69%****	0.196
CP	-0.42%	-0.05%	-0.31%	**0.05%**	-0.29%	0.54%*	-0.20%	**0.73%****	0.177
LN	0.27%	0.21%	**0.27%**	0.26%	1.19%***	**1.45%*****	1.15%***	1.37%***	3.108***
$FB+CP+LN$	0.22%	0.34%	0.29%	**0.52%****	1.16%***	1.57%***	1.20%***	**1.79%*****	2.257***
	Panel B.3: 4 years				Panel B.4: 5 years				
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV	SV
FB	0.84%**	1.12%**	0.94%**	**1.35%****	1.15%**	1.74%**	1.14%**	**1.79%****	0.647**
CP	-0.08%	0.65%*	-0.01%	**0.90%****	0.16%	**0.87%****	0.17%	0.84%*	0.375
LN	1.49%***	2.13%***	1.47%***	**2.18%*****	1.51%***	2.42%***	1.53%***	**2.48%*****	3.082***
$FB+CP+LN$	1.81%***	**2.37%*****	1.80%***	2.37%***	1.93%***	**2.92%*****	1.90%***	2.83%***	2.713***

This table reports annualized certainty equivalent return values for portfolio decisions based on recursive out-of-sample forecasts of bond excess returns for an investor with power utility (Panel A) / mean-variance utility (Panel B) and coefficient of relative risk aversion of 5. In the univariate asset allocation exercise the investor selects 2, 3, 4, *or* 5-year bond and 1-month T-bills based on the predictive density implied by a given model. In the joint asset allocation exercise the investor selects 2, 3, 4, 5-year bond and 1-month T-bills. The four 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 the combination 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 1990-2015. * significance at 10% level; ** significance at 5% level; *** significance at 1% level. For every model and maturity, we denote in bold font the CER of the estimation method (LIN, SV, TVP and TVPSV) that delivers the best result in the univariate asset allocation exercise.

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	2.58%	0.51%	3.14%	0.11%	7.20%	**8.99%****	8.53%	3.38%
CP	1.86%	**2.57%***	1.97%	1.80%*	6.13%	5.35%**	6.71%	3.87%**
LN	1.78%	**8.77%****	1.95%	**5.43%****	4.83%	**15.14%****	6.21%	**9.11%****
$FB+CP+LN$	5.37%	**10.41%****	5.51%	**7.10%****	12.84%	**23.90%****	13.32%	**13.66%****
				Panel B: 3 years				
FB	2.31%	0.01%	2.89%	-0.40%	4.77%	**5.29%****	5.73%	1.75%
CP	1.41%	**2.47%****	1.47%	**2.37%****	3.90%	**4.17%****	4.15%	3.57%**
LN	2.00%	**7.50%****	1.99%	**6.31%****	3.72%	**11.45%****	4.23%	**8.49%****
$FB+CP+LN$	4.68%	**8.62%****	4.76%	**7.15%****	8.94%	**17.40%****	9.20%	**11.39%****
				Panel C: 4 years				
FB	2.11%	0.14%	2.56%	-0.14%	3.80%	3.31%*	4.46%	1.34%
CP	1.17%	**2.47%****	1.25%	**2.52%****	3.02%	**3.81%****	3.06%	**3.49%****
LN	1.87%	**6.72%****	1.84%	**6.11%****	3.06%	**9.82%****	3.26%	**8.13%****
$FB+CP+LN$	4.05%	**7.82%****	4.09%	**7.06%****	7.11%	**14.16%****	6.98%	**10.24%****
				Panel D: 5 years				
FB	1.95%	0.42%	2.31%	0.22%	3.14%	2.49%	3.61%	1.33%
CP	1.09%	**2.67%****	1.14%	**2.66%****	2.47%	**3.71%****	2.50%	**3.38%****
LN	1.65%	**6.00%****	1.64%	**5.69%****	2.58%	**8.76%****	2.77%	**7.59%****
$FB+CP+LN$	3.54%	**7.32%****	3.57%	**6.86%****	5.84%	**11.98%****	5.67%	**9.55%****

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 the combination 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 $R_{i,0}^2 = 1 - \frac{e_{i,0}'e_{i,0}}{e_{EH,0}'e_{EH,0}}$ 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: $R_{i,1}^2 = 1 - \frac{e_{i,1}'e_{i,1}}{e_{EH,1}'e_{EH,1}}$. 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. For each model and estimation method (LIN, SV, TVP and TVPSV) we denote in bold font the R^2 in recessions which are higher than the R^2 in expansions.

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

Model	LIN Exp	LIN Rec	SV Exp	SV Rec	TVP Exp	TVP Rec	TVPSV Exp	TVPSV Rec
				Panel A: 2 years				
FB	0.48	0.43	0.46	0.32	0.47	**0.58**	0.48	0.45
CP	0.45	**0.62**	0.43	**0.47**	0.43	**0.71**	0.44	**0.61**
LN	0.38	**1.15**	0.30	**0.70**	0.38	**1.27**	0.33	**0.92**
$FB+CP+LN$	0.38	**1.21**	0.39	**0.78**	0.36	**1.42**	0.41	**1.01**
				Panel B: 3 years				
FB	0.42	0.36	0.47	0.33	0.43	**0.44**	0.48	0.39
CP	0.40	**0.54**	0.46	**0.54**	0.41	**0.60**	0.46	**0.61**
LN	0.34	**0.96**	0.36	**0.87**	0.34	**1.03**	0.37	**0.97**
$FB+CP+LN$	0.34	**0.98**	0.41	**0.90**	0.34	**1.11**	0.42	**1.01**
				Panel C: 4 years				
FB	0.39	0.34	0.48	0.35	0.39	0.37	0.48	0.38
CP	0.37	**0.48**	0.46	**0.54**	0.37	**0.52**	0.47	**0.59**
LN	0.32	**0.83**	0.38	**0.89**	0.32	**0.89**	0.38	**0.97**
$FB+CP+LN$	0.32	**0.83**	0.42	**0.92**	0.32	**0.94**	0.43	**1.00**
				Panel D: 5 years				
FB	0.36	0.32	0.47	0.35	0.36	0.35	0.48	0.38
CP	0.34	**0.45**	0.47	**0.54**	0.34	**0.48**	0.47	**0.58**
LN	0.30	**0.73**	0.40	**0.88**	0.30	**0.79**	0.40	**0.95**
$FB+CP+LN$	0.30	**0.73**	0.43	**0.90**	0.30	**0.80**	0.43	**0.96**

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 the combination 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 ($TVPSV$). For each model and estimation method (LIN, SV, TVP and TVPSV) we denote in bold font the Sharpe Ratios in recessions which are higher than their counterparts in expansions.

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

	Expected Excess Returns							
	Panel A: GDP Growth				Panel B: Consumption Growth Uncertainty			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.16	0.17*	0.09	0.09	0.08	0.08	0.09	0.10
CP	-0.29***	-0.26***	-0.33***	-0.35***	0.16	0.17*	0.17*	0.18*
LN	-0.60***	-0.59***	-0.61***	-0.61***	0.25**	0.28***	0.24**	0.28***
$FB+CP+LN$	-0.47***	-0.39***	-0.50***	-0.44***	0.26***	0.28***	0.26***	0.28***
	Panel C: Inflation Uncertainty				Panel D: Giacoletti et. al. (2016)			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.03	-0.00	0.05	0.04	0.39***	0.42***	0.43***	0.44***
CP	0.36***	0.35***	0.37***	0.34***	0.21***	0.19***	0.21***	0.19***
LN	0.48***	0.48***	0.47***	0.45***	0.09	0.09	0.10*	0.11*
$FB+CP+LN$	0.44***	0.40***	0.44***	0.39***	0.27***	0.32***	0.27***	0.33***

	Realized Utilities							
	Panel E: GDP growth Uncertainty							
	Power Utility				Mean Variance Utility			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.18*	0.19**	0.19*	0.19**	0.19*	0.19**	0.19*	0.19**
CP	0.18*	0.17*	0.18*	0.18*	0.18*	0.17*	0.18*	0.18*
LN	0.10	0.09	0.11	0.12	0.11	0.10	0.11	0.11
$FB+CP+LN$	0.11	0.10	0.11	0.11	0.11	0.10	0.11	0.12
	Panel F: Inflation Uncertainty							
	Power Utility				Mean Variance Utility			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.21**	0.22**	0.24**	0.24**	0.22**	0.22**	0.23**	0.24**
CP	0.21**	0.23**	0.21**	0.23**	0.21**	0.22**	0.21**	0.23**
LN	0.20**	0.20**	0.20**	0.22**	0.20**	0.20**	0.20**	0.22**
$FB+CP+LN$	0.19*	0.20**	0.20**	0.22**	0.19*	0.20**	0.20**	0.22**

This table reports in Panel A, B, C and D the contemporaneous correlations between out-of-sample forecasts of excess returns on a two-year Treasury bond and real GDP growth (Panel A), Consumption (Panel B) or inflation uncertainty (Panel C) and the out-of-sample bond return forecasts of Giacoletti et al. (2016) (Panel D). Panel E and F display the contemporaneous correlations between out-of-sample realized utility and GDP growth (Panel E) or inflation uncertainty (Panel F). Real GDP growth is computed as $\Delta log(GDP_t)$ where GDP_t is the real gross domestic product (GDPMC1 Fred mnemonic). 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 the combination 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. **Multivariate ICAPM test of Bali (2008)**
Panel A: ICAPM Estimates

	β (t-stat)	γ (t-stat)	δ (t-stat)	θ (t-stat)	$\beta,\gamma,\delta,\theta$ (t-stat)
Market	0.393 (0.972)	0.297 (0.763)	0.339 (0.416)	0.416 (0.992)	0.387 (0.980)
Inflation		-9.862 (-6.179)			-7.419 (-4.685)
Default			-2.179 (-4.367)		-1.640 (-3.132)
Term				-0.20 (-0.15)	1.227 (0.957)

Panel B: Correlation between expected excess returns and ICAPM risk premia

	Panel B.1: 2 years				**Panel B.2: 3 years**			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.01	0.10	0.06	0.05	0.22***	0.19***	0.23***	0.20***
CP	0.25***	0.27***	0.29***	0.34***	0.20***	0.21***	0.22***	0.22***
LN	0.29***	0.31***	0.30***	0.35***	0.15***	0.14**	0.16***	0.17***
$FB+CP+LN$	0.29***	0.28***	0.31***	0.32***	0.19***	0.19***	0.20***	0.20***

	Panel B.3: 4 years				**Panel B.4: 5 years**			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	0.13**	0.10*	0.14**	0.13**	0.27***	0.26***	0.27***	0.25***
CP	0.21***	0.21***	0.23***	0.24***	0.20***	0.21***	0.21***	0.22***
LN	0.19***	0.20***	0.20***	0.23***	0.14**	0.14**	0.15***	0.15***
$FB+CP+LN$	0.21***	0.21***	0.23***	0.23***	0.20***	0.20***	0.21***	0.21***

Panel A displays slope estimates and associated t-statistics from the following Seemingly Unrelated Regressions:

$$Rx_{i,t+1} = \alpha_i + \beta * cov(Rx_{i,t+1}, Mkt_{t+1}) + \gamma * cov(Rx_{i,t+1}, \Delta Infl_{t+1})$$
$$+ \delta * cov(Rx_{i,t+1}, \Delta Dflt_{t+1}) + \theta * cov(Rx_{i,t+1}, \Delta Term_{t+1}) + \epsilon_{i,t+1}$$

where Rx denotes the excess return of bond i, Mkt is the value-weighted excess return of the stocks belonging to NYSE, AMEX or NASDAQ, $Inflation$ is the inflation computed from the consumer price index ($CPIAUCSL$ Fred mnemonic), $Dflt$ is the difference between the BAA and AAA yields, $Term$ is the term spread computed as the difference between the long- and short-term yields; and Δ is the first-difference operator. The system contains four equations corresponding to bonds with maturity of 2, 3, 4 and 5 years. Conditional covariances are computed with the DCC (Dynamic Conditional Correlation) model of Engle (2002). As in Bali and Engle (2010) and Bali (2008) the slope coefficients β, γ, δ and θ, are pooled across equations while the intercepts (α_i) differ across equations. The t-stats are adjusted for heteroskedasticity and autocorrelation for each series and for cross-correlations across bonds based on the procedure of Parks (1967). The estimates are based on data from January 1962 to December 2015. Panel B reports the contemporaneous correlations between out-of-sample forecasts of excess returns and the risk-premia implied by the ICAPM. 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 the combination 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. *significance at 10% level; ** significance at 5% level; *** significance at 1% level.

Table 10. **Expected bond excess returns, expected utility and survey forecasts of bond yields**

	Panel A: 2 years							
	Panel A.1: Expected Bond Excess Returns							
	Slope coefficient				R^2			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-0.47***	-0.41**	-0.45***	-0.40**	4.19	1.78	4.70	2.08
CP	-0.74***	-0.60***	-0.78***	-0.61***	5.75	4.29	6.72	4.63
LN	-1.77***	-1.22***	-1.64***	-1.01***	8.64	5.32	8.04	3.69
$FB+CP+LN$	-2.03***	-1.39***	-1.80***	-1.15***	9.01	5.17	8.06	3.70

	Panel A.2: Expected Utility							
	Slope coefficient				R^2			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-0.57***	-0.49***	-0.54***	-0.47***	6.14	2.65	7.07	3.14
CP	-0.83***	-0.68***	-0.87***	-0.69***	7.29	5.67	8.38	6.06
LN	-1.87***	-1.30***	-1.73***	-1.08***	9.53	6.00	8.97	4.30
$FB+CP+LN$	-2.12***	-1.46***	-1.89***	-1.22***	9.87	5.82	8.93	4.26

	Panel B: 5 years							
	Panel B.1: Expected Bond Excess Returns							
	Slope coefficient				R^2			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-1.69***	-1.82***	-1.66***	-1.77***	12.34	10.96	11.85	11.95
CP	-1.18***	-1.05***	-1.23***	-1.14***	12.78	10.71	14.17	12.83
LN	-2.39***	-2.16***	-2.37***	-2.08***	14.56	13.30	14.15	11.57
$FB+CP+LN$	-3.17***	-2.92***	-3.07***	-2.77***	18.13	16.72	17.32	14.62

	Panel B.2: Expected Utility							
	Slope coefficient				R^2			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
FB	-1.85***	-1.90***	-1.83***	-1.85***	14.39	14.08	14.33	15.56
CP	-1.34***	-1.14***	-1.39***	-1.22***	16.24	13.05	17.74	15.46
LN	-2.56***	-2.24***	-2.53***	-2.17***	16.41	14.33	16.02	12.64
$FB+CP+LN$	-3.34***	-3.01***	-3.23***	-2.85***	19.96	18.12	19.15	15.92

This table reports the R^2 and OLS estimates of slope coefficients from a regression of the predicted bond excess return (Panels A.1 and B.1) and expected utility (Panels A.2 and B.2) on yield forecasts from the Blue Chip Financial Forecasts. 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 the combination of these. Panels A and B display results for 2- and 5-year bond maturities, respectively. 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 ($TVPSV$). All results are based on the sample 1990-2015. 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 11. **Risk-premium regression**

	Panel A: 2 years				Panel B: 3 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
β	0.35	0.65	0.37	0.73	0.19	0.37	0.19	0.43
t-stat	1.86	3.87	2.11	4.48	1.17	2.39	1.19	2.72

	Panel C: 4 years				Panel D: 5 years			
Model	LIN	SV	TVP	TVPSV	LIN	SV	TVP	TVPSV
β	0.14	0.25	0.15	0.29	0.12	0.19	0.12	0.22
t-stat	0.94	1.69	0.97	1.89	0.86	1.39	0.83	1.49

This table reports OLS estimates of the slope coefficients (and the relative t-stats) from the following regression

$$\overline{rx}_t = \mu + \beta \overline{rp}_t + u_t,$$

where $\overline{rx}_t$ denotes the predicted bond excess returns and $\overline{rp}_t$ denotes the risk premium estimates, obtained from a term structure model with unspanned macro risks, based on the approach of Joslin et. a. (2011) and Wright (2011). 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 ($TVPSV$). All results are based on the sample 1990-2015. 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 12. **Economic and statistical performance of forecast combinations**

Method	2 years	3 years	4 years	5 years
	Panel A: Out-of-sample R^2			
OW	**4.70%*****	**4.79%*****	**4.07%*****	**3.79%*****
EW	**5.53%*****	**4.33%*****	**3.49%*****	**3.12%*****
BMA	**5.53%*****	**4.61%*****	**3.78%*****	**3.53%****
	Panel B: Predictive Likelihood			
OW	**0.33*****	**0.20*****	**0.13*****	**0.10*****
EW	**0.19*****	**0.12*****	**0.08*****	**0.06*****
BMA	**0.32*****	**0.20*****	**0.13*****	**0.09*****
	Panel C: CER			
OW	**0.43%**	**1.51%*****	**2.40%*****	**2.96%*****
EW	**0.44%****	**1.30%*****	**1.83%*****	**1.91%*****
BMA	**0.50%****	**1.65%*****	**2.30%*****	**2.75%*****

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 28 forecasting models based on all possible combinations of the CP, FB and LN factors estimated using linear, SV, TVP and TVPSV 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 are the annualized certainty equivalent returns derived for an investor with power utility and a coefficient of relative risk aversion of 5 who uses the posterior predictive density implied by the forecast combination. The forecast evaluation sample is 1990:01-2015:12. * significance at 10% level; ** significance at 5% level; *** significance at 1% level.