Parameter Learning, Sequential Model Selection, and Bond Return Predictability

Andras Fulop* Junye Li† Runqing Wan‡

Amundi Working Paper

This Version: February 2017

Abstract

The paper finds statistically and economically significant evidence of out-of-sample bond return predictability for a Bayesian investor who learns about parameters, hidden states, and models over time. Both out-of-sample R2's and certainty equivalent returns suggest that predictability improves with respect to bond maturities and that introduction of stochastic volatility in a model can enhance the predictive performance to a large extent. Furthermore, we find that the factor extracted from a large panel of macroeconomic variables contains rich information on future excess bond returns. We also document that model combinations work well in forecasting excess bond returns.

Keywords: Bond Return Predictability, Bayesian Learning, Parameter uncertainty, Model Combinations.

JEL Classification: C11, G11, G12, G17

*ESSEC Business School, Paris-Singapore. Email: fulop@essec.fr.
†ESSEC Business School, Paris-Singapore. Email: li@essec.edu
‡ESSEC Business School, Paris-Singapore. Email: runqing.wan@essec.edu

1. Introduction

The expectation hypothesis of the term structure of interest rates asserts that the expected one-period return on bonds is equal to the one-period interest rate plus a risk premium, which is constant over time. A large number of empirical works have explored whether risk premia on treasury bonds are indeed constant. A standard way to test the expectations hypothesis is to run predictability regressions of excess bond returns on some predetermined predictors. Empirical investigations have uncovered some evidence of bond return predictability. Fama and Bliss (1987) and Campbell and Shiller (1991) find that excess bond returns are predictable by forward spreads or yield spreads. Cochrane and Piazzesi (2005) find that information contained in the entire term structure of interest rates can capture more than 30% of the variation of excess bond returns over the period from January 1964 to December 2003. Recently, empirical studies by Ludvigson and Ng (2009), Cooper and Priestly (2009), Huang and Shi (2014), and Joslin, Priebsch, and Singleton (2014) show that macroeconomic variables contain rich information on future excess bond returns beyond information contained in yield curve.

However, most evidence found in the above studies is in-sample. Investors in markets may be more concerned about whether there exists out-of-sample evidence of bond return predictability and whether such out-of-sample statistical predictability can translate into their economic gains. Thornton and Valente (2012) find that information contained in forward rates can not generate systematic economic value to an investor who has mean-variance preferences. Sarno, Schneider, and Wagner (2016) find under affine term structure model framework that the evident statistical predictability of bond risk premia rarely turns into investors' economic gain. However, in a recent study by Gargano, Pettenuzzo, and Timmermann (2016), by using non-overlapping excess bond returns and Bayesian MCMC methods, they find that statistically significant out-of-sample predictability can translate into economic value for a real-time investor.

In this paper, we revisit this seemingly contentious issue. We consider a Bayesian investor who faces the same learning problems as confronted by the econometrician. Except the expectations hypothesis that takes the historical mean as the optimal forecast, she has access to additional predictive models that may feature stochastic volatility. She takes parameters, state variables, and models as unknowns and updates her beliefs using Bayes rule sequentially at each time when new information becomes available. Our Bayesian investor computes the predictive return distribution at each time based on what she has learned and maximizes her expected utility by taking into account all relevant uncertainties. We implement Bayesian learning on predictive models by following the marginalized resample-move approach proposed by Fulop and Li (2013). One of key implications from Bayesian learning is that it generates persistent and long-term changes to the investor's beliefs. Our treatment here is similar to that of Johannes, Korteweg, and Polson (2014) who investigate effects of sequential learning on stock return predictability.

We take forward spreads of Fama and Bliss (1987), the forward factor proposed by Cochrane and Piazzesi (2005), and the macro factor proposed by Ludvigson and Ng (2009) as three predictors and refer to them as FB, CP, and LN. The predictive models are built by using any non-empty subsets of these three predictors. Furthermore, we consider both constant and stochastic volatility cases. Putting together, there are in total 14 individual predictive models to be considered. We also consider four model combination schemes based on boh statistical and economic evidence.

We construct monthly yields on US zero-coupon bonds with maturity 1-month, 2-, 3-, 4-, and 5-year using the updated dataset of Gurkaynak, Sack, and Wright (2007). Most studies in bond return predictability focus on predictive regressions for annual excess bond returns in monthly data. Bauer and Hamilton (2017) argue that the overlapping bond returns induce strong serial correlations in the error terms and may raise additional econometric problems when predictors are persistent. Therefore, similar to Gargano, Pettenuzzo, and Timmermann (2016), we consider one-month holding period and construct non-overlapping monthly excess bond returns. Our data range from January, 1962 to December, 2011, in total 600 months. We take the second half of the sample as the out-of-sample period, i.e., the out-of-sample period starts in January 1987 and ends in December 2011.

We explore statistical evidence of out-of-sample bond return predictability. Based on out-of-sample R^2, R^2_{OS}, in the sense of Campbell and Thompson (2008), we find statistically significant out-of-sample predictability: predictability improves with respect to bond maturities, and introduction of stochastic volatility boosts predictive performance of the models to a large extent. We also find that the models that take LN as a predictor alway perform better than the other models, no matter whether stochastic volatility is present or not. This result indicates that the macro factor extracted from a large panel of macroeconomic variables contains rich information on future excess bond returns.

We further investigate whether statistical evidence of bond return predictability eventually translates into investor's economic gains. Our investor is Bayesian. At each time, she maximizes her expected power utility over the one-period value of a portfolio consisting of a risk-free asset and a treasury bond by accounting for all sources of uncertainty. Given these portfolios, we compute certainty equivalence returns (CER) for each model to measure its predictive performance. Our results based on CER's are consistent with those based on R^2_{OS}, indicating that statistical predictability does translate into economic gains to our Bayesian investor.

The performance of model combinations is in general remarkable both statistically and economically. All the four schemes generate positive and statistically significant R^2_{OS}, which can always turn into positive certainty equivalent returns for our Bayesian investor. Furthermore, we document that a simple CER-based model combination works quite well in forecasting long-term excess bond returns.

The above results are robust to investors with different risk aversion, to the choice of different out-of-sample periods, and to the investor's initial beliefs.

Our work makes two contributions to literature. First, we provide a true real-time

Bayesian learning about bond return predictability that simultaneously takes into account belief-updating, parameter uncertainty, and model risk. Thornton and Valente (2012) and Sarno, Schneider, and Wagner (2016) follow classical approaches and therefore ignore parameter and model uncertainties. Gargano, Pettenuzzo, and Timmermann (2016) employ Bayesian MCMC methods and do allow for parameter and model uncertainties. However, to investigate out-of-sample predictability, MCMC needs to be repeatedly run at each time, leading to a large computational cost. Furthermore, the convergence of MCMC is hard to evaluate when sample size is small and stochastic volatility is introduced. Second, we provide novel results that statistical predictability of excess bond returns can translate into economic gains for a Bayesian investor who learns parameters, hidden states, and models over time with respect to information accumulation.

The remainder of the paper is organized as follows. Section 2 presents the predictive models to be considered and introduces Bayesian learning approach. Section 3 discusses how to statistically and economically evaluate predictive performance of each model. Section 4 provides data and main empirical results. Section 5 implements several robustness checks. And Section 6 concludes the paper.

2. Bayesian Learning and Bond Return Predictability

2.1. Predictive Models

In line with the existing literature, we define the log-yield of an n-year bond as

$$y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)}, \tag{1}$$

where $p_t^{(n)} = \ln P_t^{(n)}$, and $P_t^{(n)}$ is the nominal price of an n-year zero-coupon bond at time t. A forward rate with maturity n is defined as

$$f_t^{(n)} \equiv p_t^{(n-h)} - p_t^{(n)}, \tag{2}$$

and the excess return of an n-year bond is computed as the difference between the holding period return from buying an n-year bond at time t and selling it h-year later and the yield on a h-year bond at time t,

$$rx_{t+h}^{(n)} = p_{t+h}^{(n-h)} - p_t^{(n)} - h \cdot y_t^{(h)}, \tag{3}$$

where h is the holding period in years. When h takes one year, we obtain the usually used one-year holding period excess bond returns.

The standard approach to investigate bond return predictability usually takes a model of the form

$$rx_{t+1}^{(n)} = \alpha + \beta X_t + \epsilon_{t+1}, \tag{4}$$

where X_t is a set of the pre-determined predictors, $\epsilon_t \sim N(0, \sigma_{rx}^2)$ is a mean-zero constant variance error term, and the coefficients α, β, and σ_{rx} are unknown fixed parameters.

However, there is considerable evidence that suggests the bond return volatility is time-varying (Gray, 1996; Bekaert, Hodrick, and Marshall, 1997; Bekaert and Hodrick, 2001). Therefore, except the standard model (4), we also introduce the stochastic volatility model, which takes the form of

$$rx_{t+1}^{(n)} = \alpha + \beta X_t + e^{h_{t+1}} \epsilon_{t+1}, \tag{5}$$

where $\epsilon_t \sim N(0, 1)$ is a standard normal noise, and h_{t+1} is the log-volatility at time $t+1$, which is assumed to follow

$$h_{t+1} = \mu + \phi h_t + v_{t+1}, \tag{6}$$

where h_t is stationary and mean-reverting when $|\phi| < 1$, and $v_t \sim N(0, \sigma_h^2)$. For simplicity, we assume independence between ϵ_t and v_t.

Empirical studies have found that forward rates or forward spreads have ability to forecast bond returns. Fama and Bliss (1987) find that the forward-spot spread has predictive power for excess bond returns and that its forecasting power increases as the forecasting horizon becomes long. Cochrane and Piazzesi (2005) show that the whole term structure of forward rates can capture more than 30% of the variation of excess bond returns over the period from January 1964 to December 2003. Joslin, Priebsch, and Singleton (2014) provide evidence that macroeconomic variables contain rich information on yields. Recently, Ludvigson and Ng (2009) extract macro factors from a large set of 132 macroeconomic variables and show that these factors have predictive power for excess bond returns.

Therefore, in this paper, we consider three predictors from Fama and Bliss (1987), Cochrane and Piazzesi (2005), and Ludvigson and Ng (2009) and refer to them as FB, CP, and LN, respectively. Specifically, FB is defined as,

$$FB_t^{(n)} = f_t^{(n)} - y_t^{(n)}. \tag{7}$$

We construct the CP factor following Cochrane and Piazzesi (2005), i.e., at each time t, average excess bond return across maturities is regressed on the one-year bond yield and the full term structure of forward rates, $\mathbf{f}_t = [y_t^{(1)}, f_t^{(2)}, f_t^{(3)}, f_t^{(4)}, f_t^{(5)}]$,

$$\overline{rx}_{t+1} = \gamma_0 + \gamma \mathbf{f}_t + e_{t+1}, \tag{8}$$

where $\overline{rx}_{t+1} = \frac{1}{4}\sum_{n=2}^{5} rx_{t+1}^{(n)}$. Then the CP factor is computed as

$$CP_t = \hat{\gamma}_0 + \hat{\gamma}\mathbf{f}_t. \tag{9}$$

Finally, the LN factor is extracted from a large set of macroeconomic variables using principal component analysis. It is a linear combination of the estimated principal components, $\hat{\mathbf{F}}_t = [\hat{F}_{1,t}, \hat{F}_{1,t}^3, \hat{F}_{3,t}, \hat{F}_{4,t}, \hat{F}_{8,t}]$,

$$LN_t = \hat{\gamma}_0 + \hat{\gamma}\hat{\mathbf{F}}_t, \tag{10}$$

where $\hat{\gamma}_0$ and $\hat{\gamma}$ are estimated in the following regression

$$\overline{rx}_{t+1} = \gamma_0 + \gamma \hat{\mathbf{F}}_t + e_{t+1}. \tag{11}$$

In the paper, we take X_t in Equations (4) and (5) as any non-empty subset of $\{FB_t, CP_t, LN_t\}$. There are in total 7 such non-empty subsets, that is, $\{FB_t\}$, $\{CP_t\}$, $\{LN_t\}$, $\{FB_t, CP_t\}$, $\{FB_t, LN_t\}$, $\{CP_t, LN_t\}$, and $\{FB_t, CP_t, LN_t\}$, suggesting that there are in total 14 models, 7 constant volatility models and 7 stochastic volatility models. We name each model using the name(s) of its predictor(s) followed by the abbreviation of constant volatility (CV) or stochastic volatility (SV). For example, a model that takes CP and LN as its predictors and assumes stochastic volatility has a name of CPLN-SV. In model (4), when $\beta = 0$, excess bond returns are not predictable, and the optimal forecast of excess bond returns is simply the historical mean. This case is in fact the expectations hypothesis, which will be taken as a benchmark for comparison with the above 14 bond return predictive models.

2.2. Bayesian Learning and Belief Updating

We assume a Bayesian investor who faces the same belief updating problem as the econometrician (Hansen, 2007). She/he simultaneously learns about parameters, latent states, and models sequentially over time when new information arrives.

For a given predictive model $\mathcal{M}_i$, there is a set of unknown static parameters, Θ, and/or a vector of the hidden state, h_t, when stochastic volatility is introduced. The observations include a time series of excess bond returns and predictors, $y_{1:t} = \{rx_s^{(n)}, X_{s-1}\}_{s=1}^t$. At each time t, Bayesian learning consists of forming the joint posterior distribution of the static parameters and the hidden state based on information available only up to time t,

$$p(h_t, \Theta | y_{1:t}, \mathcal{M}_i) = p(h_t | \Theta, y_{1:t}, \mathcal{M}_i) p(\Theta | y_{1:t}, \mathcal{M}_i), \tag{12}$$

where $p(h_t|y_{1:t}, \Theta, \mathcal{M}_i)$ solves the state filtering problem, and $p(\Theta|y_{1:t}, \mathcal{M}_i)$ addresses the parameter inference issue. Updating of investor's beliefs therefore corresponds to updating this posterior distribution.

For the linear predictive model (4), Bayesian learning is straightforward using the particle-based algorithm proposed by Chopin (2002). However, when stochastic volatility is introduced, the model becomes non-linear and state-dependent. Therefore, for the purpose of state filtering and likelihood estimation, we design an efficient particle filter whose detailed algorithm is given in Appendix A. We note that the decomposition (12) suggests a hierarchical framework for model inference and learning. At each time t, for a given set of model parameters proposed from some proposal, we can run a particle filter, which delivers the empirical distribution of the hidden states, $p(h_t|\Theta, y_{1:t}, \mathcal{M}_i)$, and the estimate of the likelihood, $p(rx_{1:t}^{(n)}|\Theta, \mathcal{M}_i)$, that can be used for parameter learning, $p(\Theta|y_{1:t}, \mathcal{M}_i) \propto p(y_{1:t}|\Theta, \mathcal{M}_i)p(\Theta|\mathcal{M}_i)$. To achieve this aim, we rely on the marginalized resample-move approach developed by Fulop and Li (2013). The key point here is that the likelihood estimate from a particle filter is unbiased (Del Moral 2004). Furthermore, in contrast to traditional Bayesian methods, our Bayesian learning approach can be easily parallelized, making it computationally fast and convenient to use in practice.

The above Bayesian learning approach provides as natural outputs the predictive distribution of excess bond returns

$$p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i) = \int p(rx_{t+1}^{(n)}|h_t, \Theta, y_{1:t}, \mathcal{M}_i)p(h_t|\Theta, y_{1:t}, \mathcal{M}_i)p(\Theta|y_{1:t}, \mathcal{M}_i)dh_t d\Theta, \quad (13)$$

and an estimate of the marginal likelihood,

$$p(rx_{1:t}^{(n)}|\mathcal{M}_i) = \prod_{s=1}^{t-1} p(rx_{s+1}^{(n)}|y_{1:s}, \mathcal{M}_i). \quad (14)$$

Both Equations (13) and (14) account for all uncertainties related to parameters and state. Equation (14) summarizes model fit over time (model learning) and can be used to construct a sequential Bayes factor for sequential model comparison. For any two models $\mathcal{M}_i$ and $\mathcal{M}_j$, the Bayes factor at time t has the following recursive formula

$$\mathcal{BF}_t \equiv \frac{p(rx_{1:t}^{(n)}|\mathcal{M}_i)}{p(rx_{1:t}^{(n)}|\mathcal{M}_j)} = \frac{p(rx_t^{(n)}|y_{1:t-1}, \mathcal{M}_i)}{p(rx_t^{(n)}|y_{1:t-1}, \mathcal{M}_j)}\mathcal{BF}_{t-1}, \tag{15}$$

which is completely out-of-sample, and can be used for sequential comparison of both nested and non-nested models.

Bayesian learning and belief updating generate persistent and long-term shocks to the agent beliefs. To see this, define $\theta_t = E[\theta|y_{1:t}]$ as the posterior mean of a parameter θ obtained using information up to time t. The iterated expectation indicates

$$E[\theta_{t+1}|y_{1:t}] = E[E[\theta|y_{1:t+1}]|y_{1:t}] = E[\theta|y_{1:t}] = \theta_t. \tag{16}$$

Therefore, θ_t is a martingale, indicating that shocks to the agent beliefs on this parameter are not only persistent but also permanent. Thus, in Bayesian learning, the agent gradually updates her beliefs that the value of a parameter is higher or lower than that previously thought and/or that a model fits the data better than the other.

The Bayesian learning process is initialized by an agent's initial beliefs or the prior distributions. We move the fixed parameters in one block using a Gaussian mixture proposal. Given that in our marginalized approach the likelihood estimate is a complicated nonlinear function of the fixed parameters, conjugate priors are not available. For parameters that have supports of real line, we assume normal distributions for the priors. However, if a parameter under consideration has a finite support, we take a truncated normal as its prior, and if a parameter under consideration needs to be positive, we take a lognormal as its prior. The hyper-parameters of the prior distributions are calibrated using a training sample, that is, an initial dataset is used to provide information on the location and scale of the parameters. The training-sample approach is a common way to generate the objective prior distributions (O'Hagan 1994). We find that the model parameters are not sensitive to the selection of the priors. Therefore, we give flat priors to the model parameters. Set one in Table 1 provides details of the selection of functional forms and hyper-parameters for the priors.

2.3. Model Combinations

It has been found that combining forecasts across models often produces a better forecast than the individual models. Timmermann (2006) points out that model combination can be viewed as a diversification strategy that improves predictive performance in the same manner that asset diversification improves portfolio performance. Model combination is an important tool to handle model uncertainty, which is prevalent in most of empirical studies. Rapach, Strauss, and Zhou (2010) and Dangle and Halling (2012) have shown that model combinations can generate much better forecasts than the individual models in forecasting stock returns. In this paper, we consider four model combination schemes for forecasting excess bond returns.

2.3.1. Sequential Best Model

Sequential best model (SBM) selects the model with the largest marginal likelihood at each time t, i.e., it gives a weight of one to the model that has the largest marginal likelihood and a weight of zero to other models,

$$p_{SBM}(rx_{t+1}^{(n)}|y_{1:t}) = \max\left\{p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i)\right\}_{i=1}^{N}, \tag{17}$$

where N is the number of models considered. The best model may change over time, suggesting that a different model may be used for forecasting at each time.

2.3.2. Equal-weighted Model Average

Equal-weighted model average (EMA) assumes equal weight on each model, that is,

$$p_{EMA}(rx_{t+1}^{(n)}|y_{1:t}) = \sum_{i=1}^{N} w_{i,t} \times p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i), \tag{18}$$

where $w_{i,t} = 1/N$ for all i and all t. One advantage of this simple scheme is that the combining weights do not need to be estimated.

2.3.3. Bayesian Model Average

It could be beneficial to determine combining weights according to model performance. Bayesian model averaging (BMA) provides a coherent mechanism for this purpose (Hoeting et al., 1999). It is a model combination approach based on the marginal likelihood of each model,

$$p_{BMA}(rx_{t+1}^{(n)}|y_{1:t}) = \sum_{i=1}^{N} w_{i,t} \times p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i), \tag{19}$$

where $w_{i,t} = p(\mathcal{M}_i|y_{1:t})$, and $p(\mathcal{M}_i|y_{1:t})$ is the posterior probability of model i,

$$p(\mathcal{M}_i|y_{1:t}) = \frac{p(y_{1:t}|\mathcal{M}_i)p(\mathcal{M}_i)}{\sum_{j=1}^{N} p(y_{1:t}|\mathcal{M}_j)p(\mathcal{M}_j)}, \tag{20}$$

in which $p(y_{1:t}|\mathcal{M}_i)$ denotes the marginal likelihood of model i, and $p(\mathcal{M}_i)$ is the prior probability of model i. In implementation, we assume equal prior probability for each model.

2.3.4. CER-based Model Average

The above model combination schemes basically use statistical evidence to construct combining weights, $w_{i,t}$. However, investors are more concerned about whether the statistical evidence of predictability could translate into real economic gains. Therefore, it is tempting to construct combining weights according to models' economic performance.

We will see in the next section that our investor is Bayesian and tries to maximize her expected utility using the predictive distribution of excess bond returns. Models' economic performance is then evaluated using the certainty equivalence returns (CER). Therefore, we propose a simple utility-based model average scheme (UMA) that constructs combing weights using CER's at each time. Specifically,

$$p_{UMA}(rx_{t+1}^{(n)}|y_{1:t}) = \sum_{i=1}^{N} w_{i,t} \times p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_j), \tag{21}$$

where $w_{i,t} = p(M_i|y_{1:t})$ and is given by

$$p(M_i|y_{1:t}) = \frac{CER_{i,t}}{\sum_{j=1}^{N} CER_{j,t}}, \tag{22}$$

in which $CER_{i,t}$ is the certainty equivalent return computed using Equation (33) for the period from the beginning date of out-of-sample to the current time t. In empirical implementation, we find that this simple scheme works quite well.

3. Assessing Out-of-Sample Performance

3.1. Statistical Evaluation

Given the predictive distribution of excess bond returns, we can compute the posterior mean to obtain the point forecast at each time t for each model or model combination. Denote this point forecast as $\widehat{rx}_{t+1}^{(n)}$ and define the sum of squared forecast errors (SSE) from initial time of the out-of-sample period to time t as

$$\widehat{SSE}(t) = \sum_{s=1}^{t} (rx_{s+1}^{(n)} - \widehat{rx}_{s+1}^{(n)})^2. \tag{23}$$

Furthermore, the expectations hypothesis states that the optimal forecast of excess bond returns is the historical mean, that is, $\overline{rx}_{t+1}^{(n)} = \frac{1}{t}\sum_{j=1}^{t} rx_t^{(n)}$. Then the corresponding SSE is given by

$$\overline{SSE}(t) = \sum_{s=1}^{t}(rx_{s+1}^{(n)} - \overline{rx}_{s+1}^{(n)})^2. \tag{24}$$

A natural measure of the model predictive performance is the out-of-sample R^2 statistics, R^2_{OS}, proposed by Campbell and Thompson (2008). The R^2_{OS} statistic is computed as

$$R^2_{OS} = 1 - \frac{\widehat{SSE}(T)}{\overline{SSE}(T)}, \tag{25}$$

where T denotes the end of the out-of-sample period. The R^2_{OS} is analogous to the standard in-sample R^2 and measures the proportional reduction in prediction errors of the forecast from the predictive model relative to the historical mean.

It is clear that when $R^2_{OS} > 0$, the predictive model statistically outperforms the expectations hypothesis. We can further test whether this outperformance is statistically significant using the statistic developed by Clark and West (2007). The Clark-West statistic adjusts the well-known Diebold and Mariano (1995) and West (1996) statistic and generates asymptotically valid inference when comparing nested model forecasts. Specifically, the Clark-West statistic is computed by firstly defining

$$g_t = (rx_t^{(n)} - \overline{rx}_t^{(n)})^2 - \left[(rx_t^{(n)} - \widehat{rx}_t^{(n)})^2 - (\overline{rx}_t^{(n)} - \widehat{rx}_t^{(n)})^2\right], \tag{26}$$

and then regressing g_t on a constant. The Clark-West statistic is the t-statistic of this constant. Clark and West (2007) show that this statistic performs very well in terms of power and size properties.

Moreover, to trace the predictive performance of each model over time, Goyal and Welch (2008) recommend to use the cumulative difference of squared forecast errors between the expectation hypothesis and a predictive model,

$$cumSSE(t) = \overline{SSE}(t) - \widehat{SSE}(t). \tag{27}$$

A positive and increasing *cumSSE* curve indicates that the predictive model always outperforms the historical mean.

3.2. Economic Value and Certainty Equivalent Returns

The statistical predictability does not necessarily translate into economic gains for investors. We consider a real-time investor who construct a portfolio consisting of a risk-free zero-coupon bond and a risky bond with maturity n and maximizes her expected utility over the next period portfolio value, W_{t+1},

$$\max_{\omega} \mathbf{E}[U(W_{t+1})|y_{1:t}, \mathcal{M}_i], \tag{28}$$

where $U(\cdot)$ represents the investor's utility function and the portfolio value evolves according to

$$W_{t+1} = (1 - \omega_t^{(n)}) \exp(r_t^f) + \omega_t^{(n)} \exp(r_t^f + rx_{t+1}^{(n)}), \tag{29}$$

where r_t^f is the risk-free rate, and $\omega_t^{(n)}$ is the portfolio weight on the risky bond with maturity n.

We assume that our investor has a power utility with the relative risk aversion controlled by γ,

$$U(W_{t+1}) \equiv U(\omega_t^{(n)}, rx_{t+1}^{(n)}) = \frac{W_{t+1}^{1-\gamma}}{1-\gamma}. \tag{30}$$

The expected utility can be computed for each model as follows,

$$\mathbf{E}[U(W_{t+1})|y_{1:t}, \mathcal{M}_i] = \int U(\omega_t^{(n)}, rx_{t+1}^{(n)}) p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i) drx_{t+1}^{(n)}, \tag{31}$$

where the predictive distribution of excess bond returns, $p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i)$, is given by Equation (13).

Our investor is Bayesian. When computing expected utility in Equaiton (31), she takes into account all relevant uncertainties. At each time, the investor choose the portfolio weight to maximize her expected utility. In our Bayesian learning, we have M particles for each variable at each time. Then the optimal weight can be obtained by

$$\hat{w}_t^{(n)} = \arg\max \frac{1}{M} \sum_{j=1}^{M} \left\{ \frac{\left[(1-\omega_t^{(n)}) \exp(r_t^f) + \omega_t^{(n)} \exp(r_t^f + rx_{t+1}^{(n),j}) \right]^{1-\gamma}}{1-\gamma} \right\}. \tag{32}$$

In empirical implementation, we bound the portfolio weight, $\omega_t^{(n)}$, at -1 and 2 as in Thornton and Valente (2012) and Sarno, Schneider, and Wagner (2016) to prevent extreme investments (Goyal and Welch, 2008; Ferreira and Santa-Clara, 2011). In fact, the above restrictions on portfolio weight allow the investor to take full short sales.

The above portfolio weight in Equation (32) is then used to compute the investor's realized utility at each time t. We denote the realized utility from a predictive model as $\hat{U}_t$ and denote the realized utility from the historical mean as $\bar{U}_t$. Then the certainty equivalence return (CER) for each predictive model is defined as a value that equates the average realized utility from the model to that from the expectations hypothesis over the forecasting period. Following Pettenuzzo, Timmermann, and Valkanov (2014), we have

$$CER = \left(\frac{\sum_{t=1}^{T} \hat{U}_t}{\sum_{t=1}^{T} \bar{U}_t} \right)^{\frac{1}{1-\gamma}} - 1. \tag{33}$$

In addition, similar to $cumSSE$, we can also construct the cumulative sum of CER over time,

$$cumCER(t) = \sum_{s=1}^{t} \log(1 + CER_s). \tag{34}$$

where CER_t captures the real-time CER of a model and is given by,

$$CER_t = \left(\hat{U}_t / \bar{U}_t \right)^{\frac{1}{1-\gamma}} - 1. \tag{35}$$

A positive and increasing $cumCER$ curve suggests that the model always economically performs better than the historical mean.

4. Empirical Results

4.1. Data

We construct monthly yields on US zero-coupon bonds with maturity 1-month, 2-, 3-, 4-, and 5-year using the updated dataset of Gurkaynak, Sack, and Wright (2007). Most studies in bond return predictability focus on predictive regressions for annual excess bond returns in monthly data, that is, $h = 1$ in Equation (3). Bauer and Hamilton (2017) argue that the overlapping bond returns induce strong serial correlations in the error terms in predictive regressions, and may raise additional econometric problems when predictors are persistent. Therefore, similar to Gargano, Pettenuzzo, and Timmermann (2016), we consider one-month holding period and construct non-overlapping monthly excess bond returns. This implies that h is equal to $1/12$ in Equations (2) and (3).

Our sample spans from January, 1962 to December, 2011. In total, there are 600 months. Table 2 presents summary statistics for full-sample excess bond returns and predictors. Panel A shows that both mean and standard deviation of the annualized monthly excess returns increase with respect to maturity. For example, the mean excess return is about 1.41% with a standard deviation of 2.97% for the 2-year bond, whereas it increases to 2.19% with a standard deviation of 6.23% for the 5-year bond. Furthermore, we notice that both skewness and kurtosis decreases with respect to maturity. For example, the skewness and kurtosis for the 2-year excess bond returns are 0.50 and 14.9, respectively, whereas they are only 0.01 and 6.58, respectively, for the 5-year returns. This suggests that the short-maturity excess bond returns are more right-skewed and more leptokurtic than the long-maturity ones. Both short- and long-maturity excess bond returns are very weakly autocorrelated, as the first-order autocorrelations range from 0.12 (5-year) to 0.17 (2-year). Figure 1 plots the time series of excess returns for 2-, 3-, 4-, and 5-year bonds. We can see that all excess returns are quite volatile during the period of 1980-1983, whereas during the period of the recent global financial crisis, return volatility is by no means comparable.

Panel B presents summary statistics for the three predictors: FB, CP, and LN. We find that (1) the FB factors are left-skewed and their kurtosis's are close to 3, whereas both CP and LN are right-skewed and their kurtosis's are larger than 3, 5.3 and 5.6, respectively; (2) the FB factors are much more persistent than both CP and LN, and LN is the least persistent with the first-order autocorrelation being about 0.43. Panel C reports correlations between predictors. The FB factors are highly correlated with each other, with the correlation ranging from 0.88 between 2- and 5-year FB's to 0.99 between 4- and 5-year FB's. The CP factor has relatively small correlations with the FB factors, ranging from 0.46 with 2-year FB to 0.50 with 5-year FB. LN is almost not correlated with FB and CP. Its correlation with FB varies from -0.09 to 0.02, and its correlation with CP is only about 0.2.

It needs to be emphasized that when we implement Bayesian learning, both CP and LN are reconstructed at each time t using the information available only up to time t in order to avoid any hindsight problems.

4.2. Parameter Learning and Sequential Model Comparison

Different from batch estimation methods, our Bayesian learning approach provides us with the whole picture of how parameter posteriors evolve over time with respect to accumulation of information for each. In this section, we focuses on a stochastic volatility model and a constant volatility model, both of which take the LN factor as their only predictor (LN-SV and LN-CV). The learning results for other models are given in the Internet Appendix. Figure 2 presents the sequential learning of the fixed parameters for LN-SV, and Figure **??** presents the sequential learning of the fixed parameters for LN-CV, on 3-year excess bond returns. For each parameter, the posterior mean (solid line) and the (5, 95)% credible interval (dashed lines) are reported.

There are a number of notable features from these two figures. First, the investor's beliefs on parameters are quite uncertain in the beginning as the (5, 95)% credible intervals are very large for all parameters. Then, as information accumulates, the credible intervals become narrower and narrower over time, and parameter uncertainty diminishes.

Second, the speed of learning is different across parameters. For the expected return parameters, α and β, learning is faster for α than for β in both LN-SV and LN-CV. It takes only several years for α to reach small credible intervals, whereas it takes more than 10 years for β to have relatively small credible intervals. This evidence is particularly striking in LN-SV. For the parameters governing volatility, μ, ϕ, and σ_h, the learning speed for σ_h is much slower than the others. Its posterior mean is slowly going up in the beginning, and then is slowly going down after around 1970. Moreover, it takes very long time for its (5, 95)% credible interval to get sufficiently narrowed.

Third, the last panel of Figure 2 presents the sequential estimates of volatility. Consistent with the investor's beliefs on parameters, her belief on volatility is quite uncertain in the beginning, whereas after a short period, she becomes quite certain on volatility dynamics, mirrored by very narrow (5, 95)% credible intervals.

Fourth, the learning process of σ_{rx} in LN-CV reveals evidence of misspecification of the constant volatility model, as its learned value slowly drifts up and reaches its highest value around 1982 when bond returns are very volatile, and then it keeps going down up to the end of the sample.

Moreover, thanks to its recursive nature, our Bayesian learning approach produces the sequential marginal likelihood at each time t for each model i as shown in Equation (14). We can then construct the sequential Bayes factors and use them for real-time model analysis and comparison.

The last panel of Figure **??** presents the sequential log Bayes factors between LN-SV and LN-CV. It gives us a richer picture of model performance over time. First, no matter which maturity is considered, when market information is scarce and the variation of excess bond returns is very small (see Figure 1) in the beginning of the sample, LN-SV performs nearly the same as LN-CV. Second, as the market information accumulates over time, the data begin strongly to favour the stochastic volatility model. Third, most of the up-moves in Bayes factors happen during market turmoils. This phenomenon is particularly striking around 1982 when all four time series of excess bond returns have high volatility and indicates that the investors mainly update their beliefs on model specifications during market turmoils. Fourth, before 1982, the stochastic volatility model performs particularly better than the constant volatility model on the 5-year excess bond returns, whereas after 1982, its performance becomes much stronger on the 2-year excess bond returns.

4.3. Predictive Performance of Individual Models

We use the second half of the sample as the out-of-sample evaluation period, that is, the out-of-sample period is from January 1987 to December 2011, in total, 300 months. At each time t, our Bayesian learning approach provides us with the full predictive density for each model, $p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i)$, based on which we take its posterior mean as the point forecast to construct R^2_{OS} and $cumSSE$ for evaluating its predictive performance.

4.3.1. Statistical Evidence

Table 3 presents R^2_{OS}'s for all the 14 models considered. Panel A is for the constant volatility models, and Panel B for the stochastic volatility models. We have the following notable findings. First, in the 2-year excess bond returns, there is significant statistical evidence of predictability in LN-CV and FBLN-CV. For example, LN-CV generates an R^2_{OS} of 2.42%, and FBLN-CV generates an R^2_{OS} of 1.07%, Both are highly statistically significant. Second, when moving to 3-, 4-, and 5-year excess bond returns, we find that all constant volatility models that take LN as a predictor generate statistically significantly positive R^2_{OS}, and in general, predictability of these models improves with respect to maturity. For example, FBCPLN-CV has an R^2_{OS} of 0.35% in the 3-year excess bond returns, whereas its R^2_{OS} increases to 1.52% in the 4-year excess bond returns and further increases to 1.93% in the 5-year excess bond returns.

Third, when stochastic volatility is introduced, the predictability of most of the models dramatically improves, especially for those models that take LN as a predictor. For example, in 3-year excess bond returns, R^2_{OS} is about 2.64% in FBLN-CV, whereas it becomes 4.58% in FBLN-SV. This suggests that stochastic volatility is an important feature of bond return dynamics. Fourth, among constant volatility models, LN-CV performs the best no matter which maturity is concerned, and among stochastic volatility models, LN-SV outperforms the other models in 2- and 3-year excess bond returns and FBLN-SV stands out in 4- and 5-year excess bond returns. This result suggests that the macro factor extracted from a large panel of macroeconomic variables contains rich information on the yield curve.

We can also check how each model performs over time in forecasting excess bond returns by taking look at $cumSSE$ recommended by Goyal and Welch (2008). Figure 4 presents $cumSSE$'s for four models: CP-CV, CP-SV, LN-CV, and LN-SV. According to R^2_{OS}'s in Table 3, the models that take CP as the predictor performs almost the worst, whereas the models that take LN as the predictor performs almost the best. We see from Figure 4 that $cumSSE$'s for LN-CV and LN-SV are always positive and increasing over time no matter which maturity is concerned, whereas $cumSSE$'s for CP-CV and CP-SV are almost negative for all the four maturities. We also see that LN-CV performs better than LN-SV for predicting 2- and 3-year excess bond returns over time before the recent financial crisis in 2008, whereas LN-CV outperforms LN-SV for predicting 4- and 5-year excess bond returns almost all over time, even after the 2008 financial crisis.

4.3.2. Economic Evidence

The statistical evidence of predictability does not necessarily translate into economic gains for investors. Our investor is Bayesian, who takes into account all relevant uncertainty when maximizing her expected utility in Equation (28). We assume that the coefficient of investor's relative risk aversion, γ, is equal to 5, and compute the corresponding certainty equivalent return (CER) for each model using formula (33).

Table 3 presents the annualized CER values for all the 14 models considered. CER's deliver almost the same implications as R^2_{OS}'s do. We do find economic gains from bond return predictability. No matter whether constant volatility models or stochastic volatility models are considered, the economic gains increase with respect to maturities. In general, introduction of stochastic volatility improves economic gains to a large extent. However, we find that among all models, LN-CV generates the largest CER's in 2-, 4-, and 5-year excess bond returns, which are about 0.63%, 2.57%, and 2.83%, respectively, whereas LN-SV produces the largest economic gain in 3-year excess bond returns, which is about 1.91%.

Figure 5 presents the cumulative sum of CER's over time for the four models: CP-CV, CP-SV, LN-CV, and LN-SV. We see that LN-CV and LN-SV perform much better than CP-CV and CP-SV over time. The cumCER's of LN-CV and LN-SV have increasing tendency over time, even though this tendency in 2-year excess bond returns is not as strong as that in the other maturities. However, the cumCER of CP-CV is alway negative and almost decreasing over time in all four time series of excess bond returns. Even though there is evidence that the cumCER of CP-SV slightly increases over time, this evidence is too weak.

4.4. Model Combinations and Predictive Performance

We now move to take a look at statistical evidence and economic gains of bond return predictability in model combinations. Table 4 presents both R^2_{OS}'s and CER's resulted from the four model combination schemes introduced previously. We have the following main findings. First, all the four combination schemes generate highly statistically significant R^2_{OS}'s and positive CER's, no matter which maturity is concerned. Statistically significant R^2_{OS} and positive CER's indicate that there exists evidence of bond return predictability and economic gains can be achieved from this predictability. Second, having compared these model combination schemes, we find that in 2-year excess bond returns, SBM performs the best according to R^2_{OS}, whereas BMA performs the best according to CER. However, in 3-, 4-, and 5-year excess bond returns, UMA outperforms the other three schemes according to both R^2_{OS} and CER. For example, UMA generates an R^2_{OS} of 4.20% and a CER of 1.50% in 3-year excess bond returns, an R^2_{OS} of 4.44% and a CER of 2.32% in 4-year excess bond returns, and an R^2_{OS} of 4.60% and a CER of 2.87% in 5-year excess bond returns.

Figure 6 presents the cumulative sum of squared errors, and Figure 7 presents the cumulative sum of CER's, for these four model combination schemes. These two figures give us a full picture how these four combinations perform over time. The $cumSSE$'s for both SBM and UMA are always positive and have stronger increasing tendency than the other two. Similar pattern can also be found in Figure 7.

5. Robustness Checks

5.1. Risk Aversion and Predictability

We assume in our main analysis in Section 4 that the coefficient of investor's relative risk aversion is equal to 5. To see how sensitive our results are to this coefficient, we also consider another two scenarios: lower risk aversion, $\gamma = 3$, and higher risk aversion, $\gamma = 10$.

Table 5 presents CER's resulted from using these two parameters of risk aversion. In general, we see that when the investor's risk aversion becomes lower, the economic gains become smaller, whereas when risk aversion becomes higher, the economic gains become larger. For example, when $\gamma = 3$, the CER generated from LN-SV is reduced from 0.47% to 0.36; and when $\gamma = 10$, the CER generated by the same model is much higher, about 1.49%. Also, as we lower the value of risk aversion, CER decreases on average. This is mainly due to the fact that with lower γ, the weights on the risky asset reach more often the upper bound, for both the EH benchmark and other predictive models, thus making it more difficult to differentiate between these models.

However, our main results remain unchanged. First, no matter whether $\gamma = 3$ or $\gamma = 10$, the models that take LN as a predictor always perform much better. For example, among constant volatility models, LN-CV generates the highest CER's in most cases except that FBLN-CV generates the highest CER when $\gamma = 10$ in 5-year excess bond returns; and among stochastic volatility models, the model with LN as the predictor again performs the best in most cases. Second, when stochastic volatility is introduced, the economic gains generally become much stronger. Third, model combinations improve economic gains in general, and among the four combination schemes, UMA performs the best in most cases, and SBM is runner-up.

5.2. Different Out-of-Sample Periods

In Section 4, we take the second half of the sample as the out-of-sample period. In this section, we explore whether our results are sensitive to the choice of out-of-sample periods. There is substantial evidence that the Federal Reserve changes its policy rule during the early 1980s. We therefore choose an out-of-sample period that starts in January 1982 and ends in December 2011. In a recent paper, Gargano, Pettenuzzo, and Timmermann (2016) set the out-of-sample period ranging from January 1990 to December 2011. We then take this period as another out-of-sample period for robustness check.

Table 6 presents R^2_{OS}'s and CER's resulted from all the models and model combinations for the out-of-sample period ranging from January 1982 to December 2011. We find that in general with comparison to results in Table 3, the overall predictive performance of the individual models and model combinations improve a lot in this period according to both R^2_{OS} and CER. For example, in 2-year excess bond returns, the FBLN-CV model generates an R^2_{OS} of 5.96% and a CER of 0.73%, whereas these two values are only 1.07% and 0.13%, respectively, in Table 3. However, our main results still hold. The models with LN as a predictor perform better than other models; introduction of stochastic volatility dramatically improves predictive performance of the models; and all model combinations generate positive and highly significant R^2_{OS}'s, which finally translate into positive economic gains.

Table 7 presents statistical and economic evidence of predictabiliy resulted from all the models and model combinations for the out-of-sample period ranging from January 1990 to December 2011. The results are very similar to what we have see in Table 3 and Table 6. We therefore conclude that our main results are robust to the choice to out-of-sample period.

5.3. Sensitivity to Priors

Our Bayesian learning is initialized by the investor's priors on model parameters. We then test whether our results are robust to alternative priors. We use the second set of priors Table 1, which assumes a normal distribution for a parameter that has support of real line, and assumes a truncated normal distribution for a parameter that has finite support. The hyper-parameters are chosen such that the priors are not informative. We assume that the coefficient of investor's relative risk aversion is equal to 5, and take the second half of the sample as the out-of-sample period.

We obtain exactly the same results as those in Table 3 and Table 4. Therefore, we conclude that our results are not sensitive to the choice of priors at all.

6. Concluding Remarks

The paper finds statistically and economically significant evidence of out-of-sample bond return predictability for a Bayesian investor who learns about parameters, hidden states, and models at each time when new information becomes available. Except the expectations hypothesis that takes the historical mean as the optimal forecast, our Bayesian investor has access to additional predictive models. We take forward spreads of Fama and Bliss (1987), the forward factor of Cochrane and Piazzesi (2005), and the macro factor of Ludvigson and Ng (2009) as our predictors and build in total 14 individual models and 4 model combination schemes.

Most studies in bond return predictability focus on predictive regressions for annual excess bond returns in monthly data. Such overlapping bond returns induce strong serial correlations in the error terms and may raise additional econometric problems when predictors are persistent (Bauer and Hamilton, 2017). Therefore, similar to Gargano, Pettenuzzo, and Timmermann (2016), we consider one-month holding period and construct non-overlapping monthly excess bond returns. Our data range from January, 1962 to December, 2011, in total 600 months. We take the second half of the sample as the out-of-sample period. Both out-of-sample R^2 and certainty equivalent returns suggest that predictability improves with respect to bond maturities and that introduction of stochastic volatility in a model can enhance its predictive performance to a large extent. Furthermore, we find that the factor extracted from a large panel of macroeconomic variables contains rich information on future excess bond returns. We also document that a simple utility-based combination scheme works well in forecasting long-term bond returns.

REFERENCES

Bauer, M. and Hamilton, J., 2017. "Robust bond risk premia". *Working paper.*

Bekaert, G. and Hodrick, R.J., 2001. "Expectations Hypotheses Tests." *Journal of Finance* 56, 1357–1394.

Bekaert, G., Hodrick, R.J., and Marshall, D.A., 1997."On biases in tests of the expectations hypothesis of the term structure of interest rates". *Journal of Financial Economics* 44, 309–348.

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

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

Chopin, N., 2002. "A sequential particle filter method for static models". *Biometrika* 89, 539–551.

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

Cochrane, J. and Piazzesi, M., 2005. "Bond risk premia". *American Economic Review*, 94, 138–160.

Cooper, I. and Priestly, R., 2009. "Time-varying risk premiums and the output gap". *Review of Financial Studies* 22, 2801–2833.

Dangle, T. and Halling, M., 2012. "Predictive regressions with time-varying coefficients". *Journal of Financial Economics* 106, 157–181.

Del Moral, P., 2004. *Feynman-Kac formulae: Genealogical and interacting particle systems with applications.* New York: Springer.

Diebold, F. and Mariano, R., 1995. "Comparing predictive accuracy". *Journal of Business & Economic Statistics* 13, 253–263.

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

Ferreira, M.A. and Santa-Clara, P., 2011. "Forecasting stock market returns: The sum of the parts is more than the whole." *Journal of Financial Economics* 100, 514–537.

Fulop, A., and Li, J., 2013. "Efficient learning via simulation: A marginalized resample-move approach". *Journal of Econometrics* 176, 146–161.

Fulop, A., Li, J., and Yu, J., 2015. "Self-Exciting Jumps, Learning, and Asset Pricing Implications." *Review of Financial Studies* 28, 876–912.

Gargano, A., Pettenuzzo, D., and Timmermann, A.G., 2017. "Bond return predictability: Economic value and links to the macroeconomy". *Working Paper.*

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

Gray, S.F., 1996. "Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process". *Journal of Financial Economics* 42, 27–62.

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

Hansen, L., 2007. "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk". *American Economic Review* 97, 1–30.

Hoeting, J., Madigan, D., Raftery, A. and Volinsky, C.,1999. "Bayesian model averaging: A tutorial". *Statistical Science* 14, 382–417.

Huang, J. and Shi, Z., 2014. "Determinants of Bond Risk Premia". *Working Paper.*

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

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

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

O'Hagan, A. 1994. *Kendall's advanced theory of statistics, vol. 2B: Bayesian inference.* New York: Halsted Press.

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

Pettenuzzo, D., Ravazzolo, F., 2016. Optimal Potfolio Choice under Decision-Based Model Combinations, Journal of Applied Econometrics 31, 1312-1332.

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

Sarno, L., Schneider, P., and Wagner, C., 2016. "The economic value of predicting bond risk premia". *Journal of Empirical Finance* 37, 247–267.

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

Timmermann, A., 2006. "Forecast combinations". In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), *Handbook of Economic Forecasting* vol. 1. Elsevier: Amsterdam.

West, K.D., 1996. "Asymptotic inference about predictive ability". *Econometrica* 64, 1067–1084.

Appendix A: Particle Filter Algorithm

Our state-space model takes the form:

$$rx_{t+1} = \alpha + \beta x_t + exp(h_{t+1})\epsilon_{rx,t+1} \quad (36)$$

$$h_{t+1} = \mu + \phi h_t + \sigma_h \epsilon_{h,t+1} \quad (37)$$

where rx_t, x_t are observations of excess returns and predictors. The first equation is the observation equation and second is the state transition equation. Because of the stochastic volatility, the model is non-linear and non-gaussian.

Denote the parameter set $\theta = \{\alpha, \beta, \mu, \phi, \sigma_h\}$. We employ a bootstrap particle filter as follows:

Step 1: We initialize the filter and set state particles to be: $h^{(i)} = \mu/(1-\phi), i = 1, ..., M$ and give each particle a weight 1/M.

Notice that there is a lag between the observation rx and the predictor x. For time t=2,...,T:

Step 2: We run the particle filter on the new observation. Run one step forward to sample new states using the transition equation to get $\psi(h_t^{(i)}|h_{t-1}^{(i)}, \theta)$. Compute the particle weight by: $w_t^{(i)} \sim Normal(rx_t - \alpha - \beta x_{t-1}, exp(h_t^{(i)}))$. Normalize the weight by $\pi_t^{(i)} = w_t^{(i)} / \sum_{k=1}^{M} w_t^k$.

Step 3: Stratified resample step: first draw i.i.d. uniform random numbers $U_t^{(i)}, i = 1, ..., M$, then draw new particle indices by inverting the cumulative distribution function (CDF) of the multinomial characterized by $\pi_t^{(i)}$ at the stratified uniforms $\frac{i+U_t^{(i)}}{M}$. Update and draw state particles population according to the new indices and the new sample is equally weighted now. The marginal likelihood can be computed as: $p(rx_t|rx_{1:t}, \theta) =$

$\frac{1}{M}\sum_{k=1}^{M} p(rx_t|rx_{1:t-1}, h_t^k)$.

Appendix B: Learning Algorithm for Stochastic Volatility Model

Our state-space model takes the form:

$$rx_{t+1} = \alpha + \beta x_t + exp(h_{t+1})\epsilon_{rx,t+1} \tag{38}$$

$$h_{t+1} = \mu + \phi h_t + \sigma_h \epsilon_{h,t+1} \tag{39}$$

where rx_t, x_t are observations of excess returns and predictors. Denote the parameter set $\theta = \{\alpha, \beta, \mu, \phi, \sigma_h\}$. The first equation is the observation equation and second is the state equation. Because of the stochastic volatility, the model is non-linear and non-gaussian.

Step 1: We initialize the algorithm by sampling N parameter particles: $\theta^{(n)} \sim Prior(\theta), k = 1, ...N$ from the prior distributions. We choose a very large prior for each parameter to make sure that it is completely unrestrictive. Then, for each parameter set $\theta^{(n)}$, we give them weight $w_0^{(n)} = 1/N$, set the initial state particles: $h_{t=0}^{(i,n)} = \mu^{(n)}/(1-\phi^{(n)}), i = 1, ..., M$ and give each particle a weight $1/M$. Bear in mind that for each parameter set $\theta^{(n)}$, we have M state particles.

Because there is a lag between the observation rx and predictor x, for time t=2,...T:

Step 2: We run the particle filter on the new observation for each $\theta^{(n)}$: we run one step forward to sample states using the transition equation above: $\psi(h_t^{(n,i)}|h_{t-1}^{(n,i)}, \theta^{(n)})$. Then we can compute the marginal likelihood on each parameter set: $\hat{p}(rx_t|rx_{1:t-1}, x_{t-1}, \theta^{(n)})$, which will be the incremental weight for each $\theta^{(n)}$.

Step 3: Compute the new weight on each $\theta^{(n)}$ as: $w_t^{(n)} = w_{t-1}^{(n)} \times \hat{p}(rx_t|rx_{1:t-1}, x_{t-1}, \theta^{(n)})$. Then we normalize the weight to get $\pi_t^{(n)} = s_t^{(n)}/\sum_{k=1}^{N} s_t^{(k)}$ and compute the effective sample size ESS as: $ESS_t = \frac{1}{\sum_{k=1}^{N}(\pi^{(k)})^2}$.

Step 4: If $ESS_t < B_1$, we perform a resample-move step. This aims to handle the famous problem of particle deterioration and will enrich the parameter sets. We choose the bound to be $\frac{1}{2} \times N$. (1). We resample the parameter sets and state particles associated according to their normalized weights $\pi_t^{(n)}$ and obtain an equally-weighted sample $\{\theta^{(n)}, h_t^{(n),i=1,...,M}, \hat{p}(rx_{1:t|\theta^{(n)}}); n = 1, ..., N\}$. (2). Then we pass the whole population through an MCMC kernel from Andrieu et al.(2010) with a proposal distribution: $\psi(\theta, h_{1:t}|\theta') = \psi_t(\theta|\theta')\psi(h_{1:t|\theta})$, where we choose the proposal distribution $\psi_t(\theta|\theta')$ to be an independent multivariate normal with its mean and covariance fitted to the posterior of θ. To measure the efficiency of the move step, the acceptance probability of the newly proposed sets:

$$min\{1; \frac{p(\theta^*)\hat{p}(rx_{1:t}|\theta)^*}{p(\theta^{(n)})\hat{p}(rx_{1:t}|\theta^{(n)})} \frac{h_t(\theta^{(n)}|\theta^*)}{h_t(\theta^*|\theta^{(n)})}\} \tag{40}$$

. The sequential Bayes factor for comparing model M_1 and M_2, can be given by the recursive formula:

$$BF_t = \frac{p(rx_{1:t}|M_1)}{p(rx_{1:t}|M_2)} = \frac{p(rx_t|y_{1:t-1}, M_1)}{p(rx_t|y_{1:t-1}, M_2)} BF_{t-1} \tag{41}$$

where $p(rx_t|rx_{1:t-1}, M_1)$ can be computed in Step 2 for each model.

Table 1: **The Prior Distributions**

	Set One	Set Two
α	N(0, 10)	N(0, 10)
β	N(0, 10)	N(0, 10)
σ_{rx}	$\log(\sigma_{rx}) \sim$ N(-2, 5)	Truncated Normal: N(0, 10), $\sigma_{rx} > 0$
μ	N(0, 5)	N(0, 5)
ϕ	Truncated Normal: N(0, 5), $\phi \in$(-1, 1)	Truncated Normal: N(0, 5), $\phi \in$(-1, 1)
σ_h	$\log(\sigma_h) \sim$ N(-2, 5)	Truncated Normal: N(0, 15), $\sigma_h > 0$

The table shows two sets of prior distributions we consider. The linear model is given in equation (4). Parameters for linear models are: $\alpha, \beta, \sigma_{rx}$ The SV model is given in equation (5) and (6). Parameters for SV models are: α, β, μ, ϕ, σ_h.

Table 2: **Summary Statistics**

Panel A: Execess Bond Returns

	2-Year	3-Year	4-Year	5-Year
Mean	1.415	1.732	1.987	2.194
St.dev	2.971	4.156	5.217	6.225
Skew	0.500	0.208	0.057	0.015
Kurt	14.86	10.65	7.900	6.580
AC(1)	0.169	0.153	0.138	0.124

Panel B: Predictors

	FB					
	2-Year	3-Year	4-Year	5-Year	CP	LN
Mean	0.108	0.129	0.145	0.158	0.153	0.153
St.dev	0.100	0.116	0.129	0.138	0.213	0.308
Skew	-0.072	-0.238	-0.223	-0.143	0.785	0.841
Kurt	3.756	3.378	3.041	2.755	5.288	5.613
AC(1)	0.878	0.899	0.913	0.923	0.673	0.428

Panel C: Correlations

	FB2	FB3	FB4	FB5	CP	LN
FB2	1.000	0.973	0.926	0.879	0.460	-0.093
FB3		1.000	0.987	0.961	0.472	-0.055
FB4			1.000	0.993	0.490	-0.016
FB5				1.000	0.500	0.017
CP					1.000	0.187
LN						1.000

This table presents the summary statistics of bond excess returns and full-sample predictors. Panel A shows the mean, standard deviation, skewness, kurtosis and first-order autocorrelation of annualized monthly excess returns (in percentage). Panel B shows the mean, standard deviation, skewness, kurtosis and first-order autocorrelation of predictors. Panel C shows the correlation matrix of predictors. Full-sample data is from Jan, 1962 to Dec, 2011.

Table 3: **Out-of-Sample Predictability**

Panel A: Constant Volatility Models								
	2-Year		3-Year		4-Year		5-Year	
	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER
FB	0.02	-0.95	1.14	-0.36	1.51**	0.28	1.66**	0.54
CP	-4.62	-1.06	-2.53	-0.96	-1.27	-0.90	-0.64	-0.79
LN	2.42***	0.63	3.73***	1.58	4.97***	2.57	4.79***	2.83
FBCP	-4.25	-0.17	-2.39	-0.81	-0.56	-0.09	-0.17*	0.07
FBLN	1.07***	0.13	2.64***	0.86	3.15***	1.68	3.31***	2.11
CPLN	-2.09***	-0.08	0.92***	0.62	2.36**	1.59	3.11**	2.08
FBCPLN	-1.71***	-0.03	0.35***	0.28	1.52***	0.81	1.93***	1.21
Panel B: Stochastic Volatility Models								
	2-Year		3-Year		4-Year		5-Year	
	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER
FB	-1.33	-1.00	0.21*	-0.67	1.25**	0.20	1.43**	0.49
CP	-0.81	0.07	-0.39	0.11	0.41	0.35	0.86	0.68
LN	5.43***	0.47	4.64***	1.91	4.21***	2.49	3.71**	2.66
FBCP	-0.63	-0.77	0.41*	-0.37	1.26**	0.48	1.85**	0.81
FBLN	5.03***	0.36	4.58***	1.16	4.53***	1.90	4.18***	2.11
CPLN	3.40***	0.46	3.34***	1.24	4.04***	2.08	4.11**	2.73
FBCPLN	4.21***	0.35	3.69***	0.96	4.03***	1.54	3.97***	2.16

This table presents the out-of-sample R-squared for linear and stochastic volatility forecasting models. Linear or Stochastic models can use different combinations of predictors: one predictor (FB, CP, LN), two predictors (FBCP, FBLN, CPLN), or three predictors (FBCPLN). The $OOS - R^2$ is given in equation (25). The statistical significance measure is from Clark and West (2007). * means significance at 10% level. ** means significance at 5% level. *** means significance at 1% level. The CER is given in equation (33). Risk aversion is 5. The out-of-sample period is from Jan, 1987 to Dec, 2011.

Table 4: **Model Combinations and Predictability**

	2-Year		3-Year		4-Year		5-Year	
	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER
SBM	5.74***	0.50	4.08***	1.16	3.39***	1.77	3.55***	2.84
EMA	4.47***	0.07	4.07***	0.70	4.31***	1.49	4.24***	2.11
BMA	5.24***	0.58	2.78***	0.15	3.06***	0.92	2.65***	0.93
UMA	4.00***	0.56	4.20***	1.50	4.44***	2.32	4.60***	2.87

The table presents out-of-sample R-squared and CER results for model combinations in the out-of-sample period. The $OOS - R^2$ is given in equation (25). The statistical significance measure is from Clark and West (2007). * means significance at 10% level. ** means significance at 5% level. *** means significance at 1% level. The CER is given in equation (33). Risk-aversion is 5. Out-of-sample period is from Jan, 1987 to Dec, 2011.

Table 5: **Out-of-Sample Predictability: Risk Aversions**

Panel A: Constant Volatility Models

	2-Year		3-Year		4-Year		5-Year	
	$\gamma = 3$	$\gamma = 10$	$\gamma = 3$	$\gamma = 10$	$\gamma = 3$	$\gamma = 10$	$\gamma = 3$	$\gamma = 10$
FB	-1.11	-0.03	-1.14	0.58	-0.50	0.93	0.07	1.03
CP	-1.05	-0.50	-1.66	-0.31	-1.52	-0.22	-1.72	-0.22
LN	0.49	1.34	0.78	1.93	1.95	2.21	2.66	2.01
FBCP	-1.12	-0.28	-1.55	-0.09	-0.90	0.31	-0.15	0.30
FBLN	0.12	1.05	-0.08	1.90	0.57	2.20	1.28	2.32
CPLN	-0.13	0.84	-0.22	1.44	0.48	1.57	1.35	1.49
FBCPLN	-0.11	0.69	-0.50	1.34	-0.21	1.75	0.33	1.79

Panel B: Stochastic Volatility Models

	2-Year		3-Year		4-Year		5-Year	
	$\gamma = 3$	$\gamma = 10$	$\gamma = 3$	$\gamma = 10$	$\gamma = 3$	$\gamma = 10$	$\gamma = 3$	$\gamma = 10$
FB	-1.19	-0.03	-1.68	0.25	-0.66	0.50	-0.08	0.44
CP	0.02	0.71	-0.56	0.40	-0.56	0.51	0.20	0.62
LN	0.36	1.49	1.10	2.34	1.90	2.33	2.42	2.02
FBCP	-0.92	0.10	-1.18	0.30	-0.32	0.63	0.44	0.68
FBLN	0.31	1.18	0.45	1.79	1.03	2.00	1.36	1.71
CPLN	0.45	1.27	0.40	1.84	1.15	2.18	2.23	2.12
FBCPLN	0.28	1.18	0.25	1.65	0.73	1.89	1.56	1.76

Panel C: Model Combinations

	2-Year		3-Year		4-Year		5-Year	
	$\gamma = 3$	$\gamma = 10$	$\gamma = 3$	$\gamma = 10$	$\gamma = 3$	$\gamma = 10$	$\gamma = 3$	$\gamma = 10$
SBM	0.33	1.47	0.35	1.92	0.86	1.68	2.28	1.80
EMA	0.15	0.94	-0.19	1.53	0.46	1.74	1.09	1.58
BMA	0.53	1.46	-0.73	1.02	0.08	1.07	0.51	0.67
UMA	0.47	1.30	0.74	2.10	1.45	2.17	2.16	1.92

The table presents a robustness check of Certainty Equivalent Returns for linear models, stochastic volatility models and model combination schemes, relative to the EH benchmark, in the out-of-sample period. The CER is given in equation (33). Risk aversion coefficient γ is 3 or 10. Out-of-sample period is from Jan, 1987 to Dec, 2011.

Table 6: **Out-of-Sample Predictability: 1982-2011**

Panel A: Constant Volatility Models

	2-Year		3-Year		4-Year		5-Year	
	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER
FB	3.87***	-0.10	3.36***	0.89	3.48***	1.79	3.45***	2.16
CP	-3.69	-1.17	-2.08	-0.89	-0.99	-0.61	-0.39	-0.46
LN	2.77***	0.65	3.48***	1.94	4.29***	3.10	3.98***	3.28
FBCP	1.17**	-0.39	1.43**	0.24	2.27**	1.10	2.51***	1.39
FBLN	5.96***	0.73	5.48***	1.87	5.61***	3.18	5.63***	3.86
CPLN	-0.18***	0.25	1.39***	1.25	2.39***	2.31	2.71***	2.63
FBCPLN	3.84***	0.69	4.31***	1.66	4.64***	2.55	4.79***	3.13

Panel B: Stochastic Volatility Models

	2-Year		3-Year		4-Year		5-Year	
	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER
FB	3.16**	-0.25	3.41***	0.34	3.73***	1.25	3.46***	1.47
CP	-1.27	-0.14	-0.30	0.20	0.28	0.51	0.69	0.85
LN	5.05***	0.37	4.09***	1.71	4.06***	2.46	3.38***	2.56
FBCP	3.32***	-0.22	3.60***	0.50	3.90***	1.44	4.12***	1.78
FBLN	9.05***	1.06	7.30***	2.12	6.79***	2.87	5.98***	2.94
CPLN	4.05***	0.56	3.97***	1.62	4.29***	2.32	3.97***	2.84
FBCPLN	8.52***	1.09	7.12***	2.11	6.65***	2.70	6.08***	3.04

Panel C: Model Combinations

	2-Year		3-Year		4-Year		5-Year	
	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER
SBM	6.62***	0.99	3.93***	1.38	3.88***	2.49	4.76***	3.71
EMA	6.41***	0.70	5.47***	1.70	5.19***	2.38	4.90***	2.84
BMA	5.13***	0.54	4.85***	0.94	4.39***	1.70	4.08***	1.70
UMA	5.27***	0.63	5.03***	1.78	4.90***	2.46	4.88***	2.88

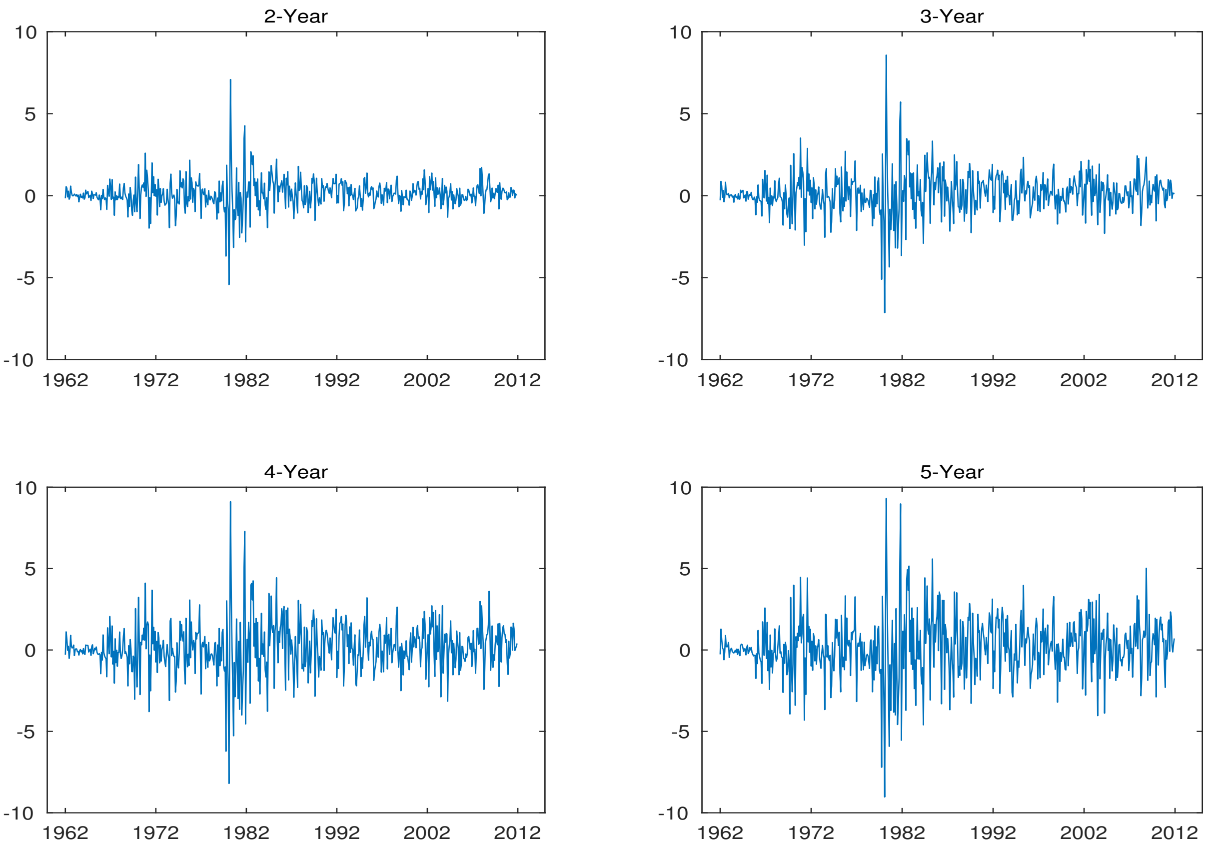
This table presents the out-of-sample R-squared and CERs for linear models,stochastic volatility forecasting models and model combinations. The R^2 is given in equation (25). The statistical significance measure is from Clark and West (2007). * means significance at 10% level. ** means significance at 5% level. *** means significance at 1% level. The CER is given in equation (33). Risk aversion is 5. The out-of-sample period is from Jan, 1982 to Dec, 2011.

Table 7: **Out-of-Sample Predictability: 1990-2011**

Panel A: Constant Volatility Models								
	2-Year		3-Year		4-Year		5-Year	
	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER
FB	0.87*	-0.79	2.23**	0.10	2.69**	0.91	2.93**	1.31
CP	-1.81	-1.08	-0.11	-0.64	0.31	-0.42	0.65	-0.30
LN	1.61***	0.33	3.78***	1.46	5.01***	2.39	5.08***	2.75
FBCP	-0.70*	-0.94	0.76*	-0.27	2.04**	0.65	2.40**	1.07
FBLN	-0.04***	0.02	2.72***	0.97	3.48***	2.00	4.00***	2.75
CPLN	-1.94***	-0.08	1.97***	0.95	3.43**	2.05	3.90**	2.46
FBCPLN	-1.59***	0.01	1.55***	0.77	2.78***	1.58	3.42***	2.23
Panel B: Stochastic Volatility Models								
	2-Year		3-Year		4-Year		5-Year	
	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER
FB	-1.14	-0.76	1.02*	-0.15	1.91**	0.90	2.31**	1.44
CP	0.24	0.12	0.68	0.61	0.93	0.79	1.32	1.05
LN	5.94***	0.32	5.24***	1.82	4.68***	2.49	4.04***	2.66
FBCP	0.31*	-0.55	1.73**	0.18	2.54**	1.20	3.14**	1.74
FBLN	5.81***	0.31	5.84***	1.72	5.47***	2.57	5.31***	3.05
CPLN	4.85***	0.54	4.76***	1.71	5.09***	2.66	5.18**	3.40
FBCPLN	5.37***	0.39	5.45***	1.53	5.51***	2.32	5.39***	3.19
Panel C: Model Combinations								
	2-Year		3-Year		4-Year		5-Year	
	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER	R^2_{OS}	CER
SBM	6.58***	0.51	4.89***	1.34	4.00**	1.92	4.04**	3.08
EMA	5.70***	0.10	5.34***	1.13	5.24***	2.01	5.28***	2.58
BMA	5.79***	0.55	4.19***	0.75	4.13**	1.73	3.67**	1.75
UMA	4.20***	0.27	4.88***	1.29	5.37***	2.37	5.51***	2.65

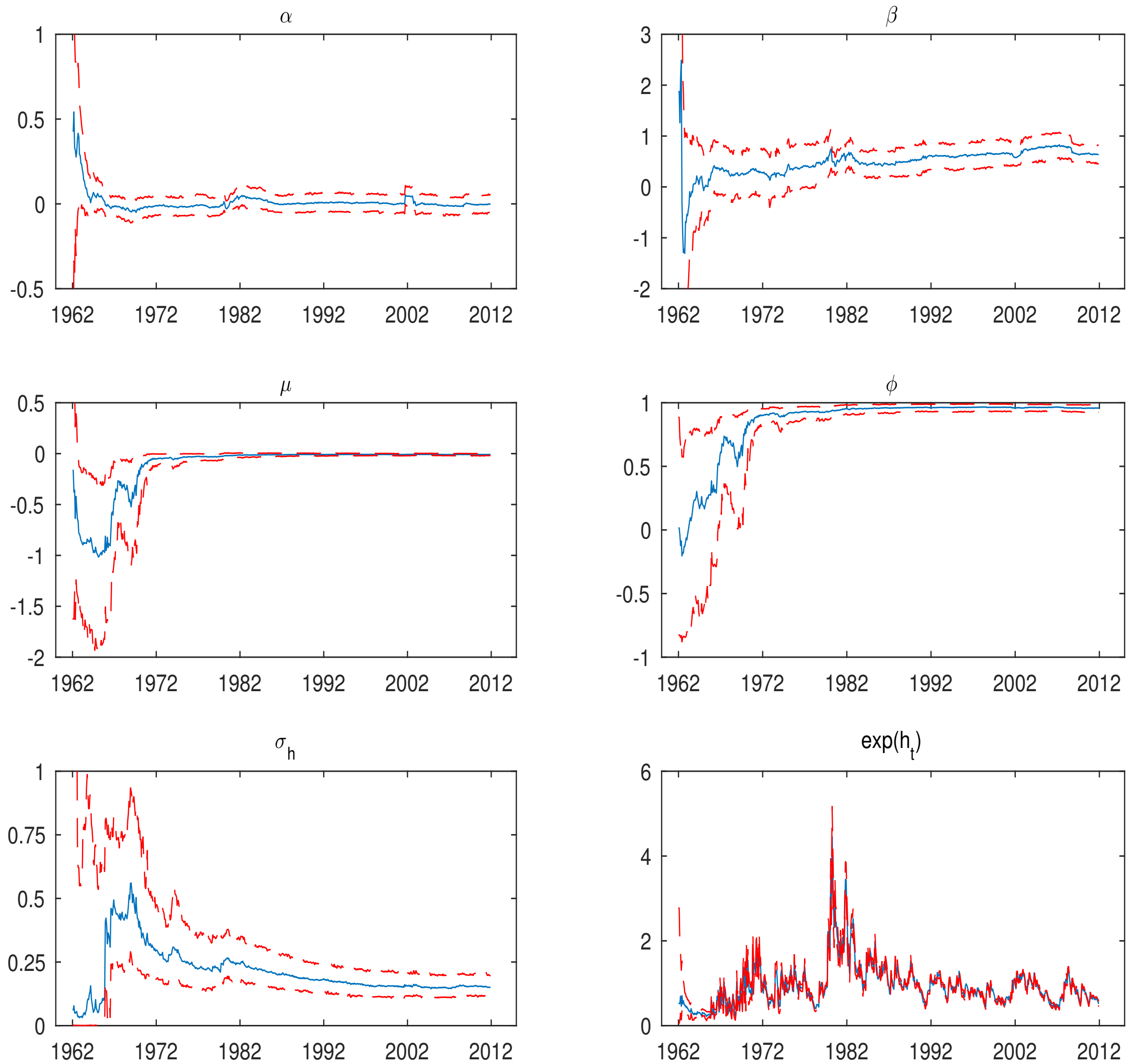
This table presents the out-of-sample R-squared and CERs for linear models,stochastic volatility forecasting models and model combinations. The R^2 is given in equation (25). The statistical significance measure is from Clark and West (2007). * means significance at 10% level. ** means significance at 5% level. *** means significance at 1% level. The CER is given in (33). Risk-aversion is 5. The out-of-sample period is from Jan, 1990 to Dec, 2011.

Figure 1: **The Time Series of Excess Bond Returns**



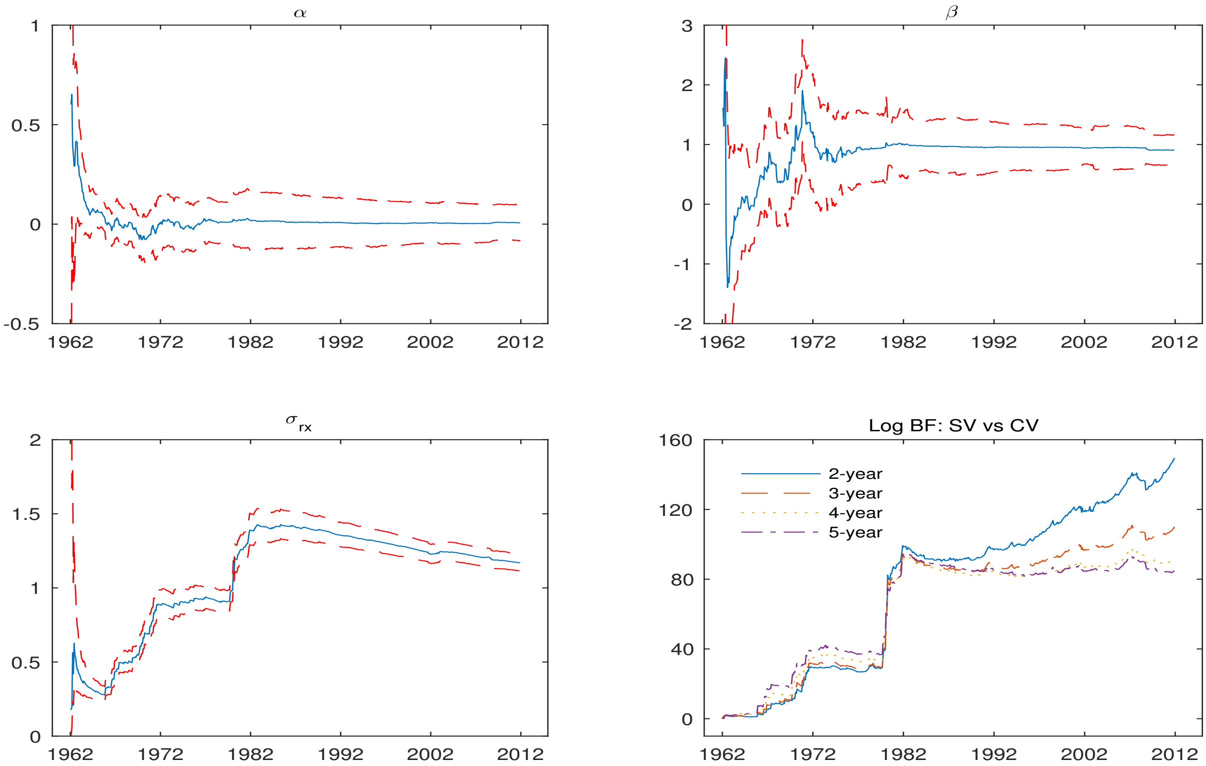
This figure plots the time series of 4 excess bond returns (in percentage), from Jan, 1962 to Dec, 2011.

Figure 2: **Parameter Learning for LN-SV**



The figure shows time series parameter estimates of stochastic volatility model with LN predictor for 3-year bond excess returns. The estimation is based on full-sample information. The model form is given in equation (5) and (6). The last panel shows the stochastic volatility estimate. The two dashed lines are 5-th and 95-th percentiles of estimate distribution. The solid line is the mean estimate for each parameter. Full-sample is from Jan, 1962 to Dec, 2011.

Figure 3: **Parameter Learning for LN-CV and Log Bayes Factors**

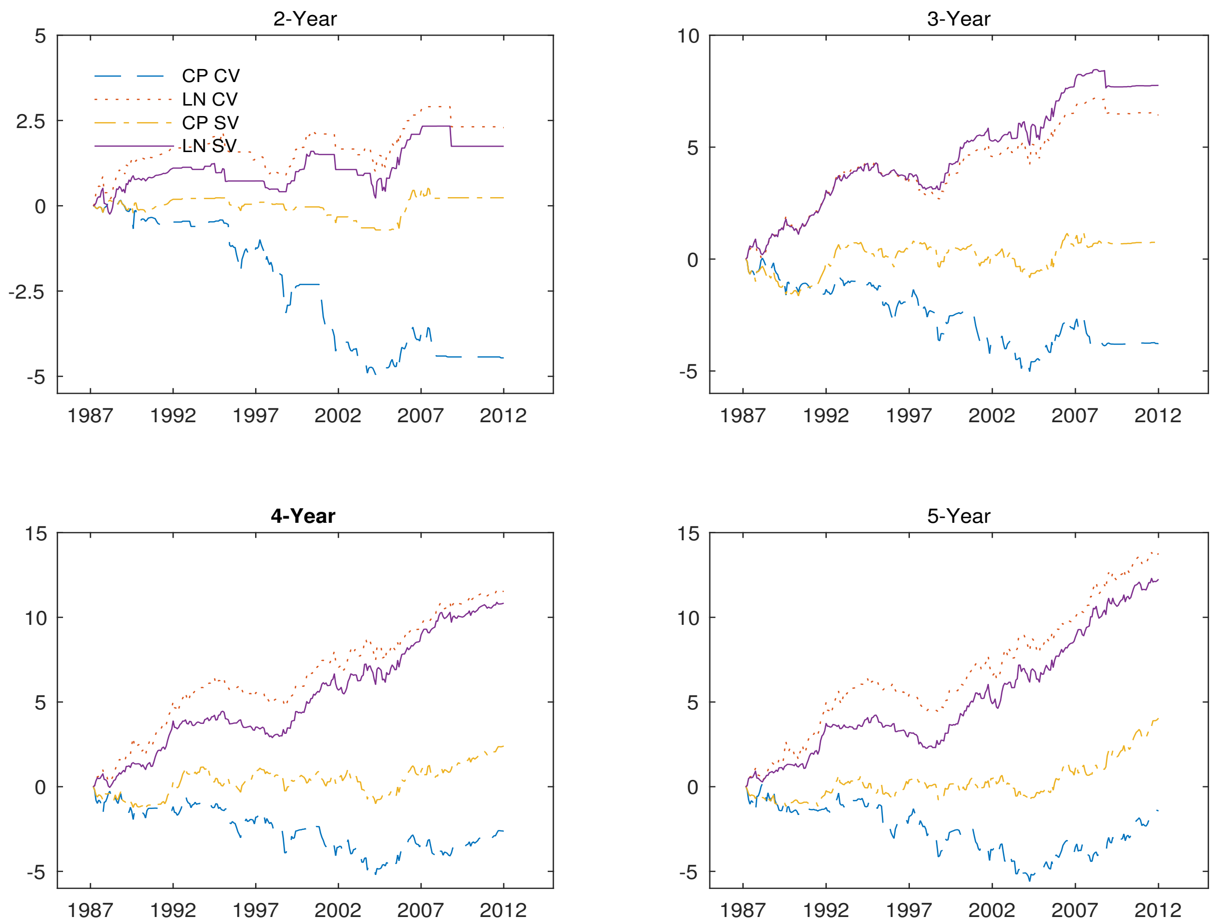


The figure shows time series parameter estimates of constant volatility model with LN predictor for 3-year bond excess returns. The estimation is based on full-sample information. The model form is given in equation (4). The last panel shows the log Bayes factor of LN-SV and LN-CV models, for all 4 maturities. The two dashed lines are 5-th and 95-th percentiles of estimate distribution. The solid line is the mean estimate for each parameter. Full-sample is from Jan, 1962 to Dec, 2011.

Figure 4: **The *cumSSE*'s for Individual Models**

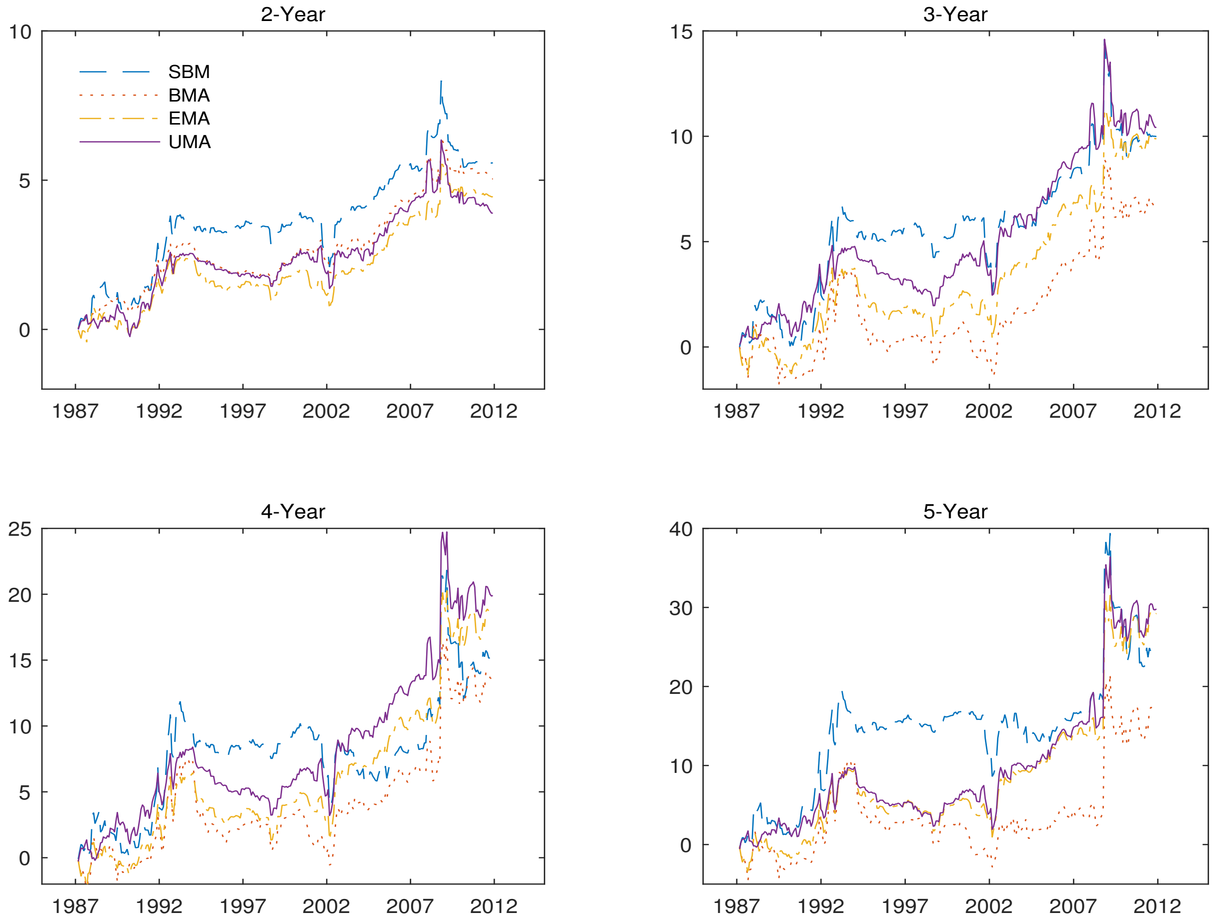
This figure shows $cumSSE$'s for CP-CV, CP-SV, LN-CV and LN-SV models. $cumSSE$ is given in equation (27). Out-of-sample period is from Jan, 1987 to Dec, 2011.

Figure 5: **The** ***cumCER*****'s for Individual Models**



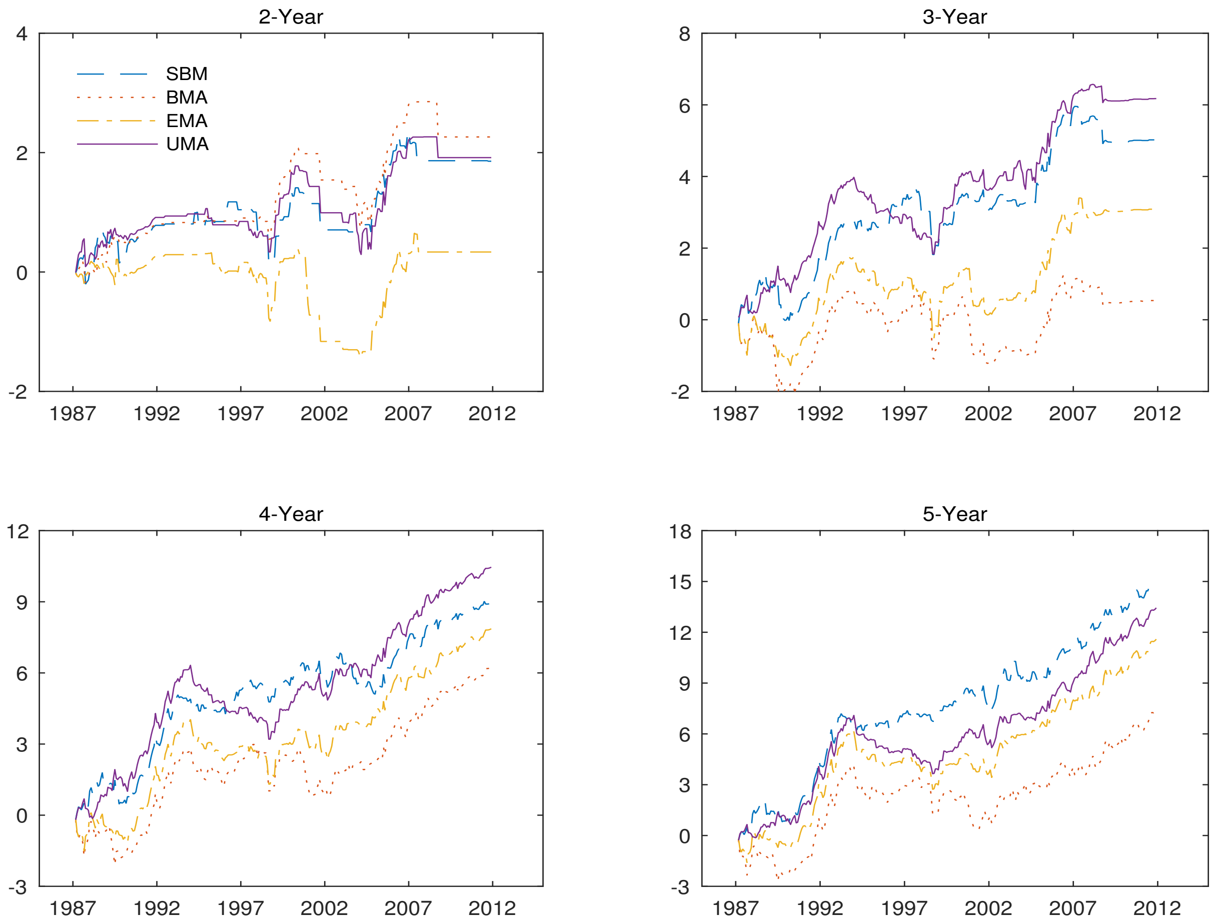
This figure shows *cumCER*'s for CP-CV, CP-SV, LN-CV and LN-SV models. *cumCER* is given in equation (34) and (35). Out-of-sample period is from Jan, 1987 to Dec, 2011.

Figure 6: **The $cumSSE$'s for Model Combinations**



This figure shows $cumSSE$'s for four model combination schemes. $cumSSE$ is given in equation (27). Out-of-sample period is from Jan, 1987 to Dec, 2011.

Figure 7: **The** $cumCER$**'s for Model Combinations**



This figure shows $cumCER$'s for four model combination schemes. $cumCER$ is given in equation (34) and (35). Out-of-sample period is from Jan, 1987 to Dec, 2011.