Real-Time Learning and Bond Return Predictability

Andras Fulop[*] Junye Li[†] Runqing Wan[‡]

This Version: January 2018

Abstract

The paper examines statistical and economic evidence of out-of-sample bond return predictability for a real-time Bayesian investor who learns about parameters, hidden states, and predictive models over time. We find some statistical evidence using information contained in forward rates. However, such statistical predictability can not generate any economic value for investors. Furthermore, statistical and economic evidence from fully revised macroeconomic data completely vanishes when real-time macroeconomic information is used. We also show that highly levered investments in bonds can improve short-run bond return predictability.

Keywords: Learning, Bond Return Predictability, Parameter Uncertainty, Model Combinations, Real-Time Macroeconomic Information.

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 expectations hypothesis (EH) 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 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. However, such evidence is statistical and 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 economic gains. Thornton and Valente (2012) find that information contained in forward rates cannot generate systematic economic value to an investor who has mean-variance preferences. Sarno, Schneider, and Wagner (2016) find that under affine term structure model framework the evident statistical predictability of bond risk premia rarely turns into investors' economic gain.

Recently, empirical studies by Ludvigson and Ng (2009), Cooper and Priestly (2009), Huang and Shi (2014), Joslin, Priebsch, and Singleton (2014) and Jiang and Tong (2017) among others show that macroeconomic variables contain rich information on future excess bond returns beyond information only in yield curve. Gargano, Pettenuzzo, and Timmermann (2017) implement an out-of-sample investigation using non-overlapping excess bond returns and Bayesian Markov Chain Monte Carlo (MCMC) methods. They find strong evidence that statistically significant out-of-sample bond return predictability by the macroeconomic factor can translate into economic value. However, nearly all these studies use the fully revised macroeconomic data and ignore issues related to data revision and publication delay. A recent paper by Ghysels, Horan, and Moench (2017) argue that macroeconomic data revision may result in spurious evidence of bond return predictability.

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 which assumes no predictability, she has access to additional predictive models that may feature stochastic volatility. She takes parameters, latent states, and/or predictive models as unknowns and updates her beliefs using Bayes' rule in real time with respect to information accumulation. 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). This algorithm is generic, efficient, and highly parallel in the sense that it does not suffer from the convergence issue and requires minor computational and design effort with comparison to traditional Bayesian MCMC methods. Our treatment here is similar to those of Johannes, Korteweg, and Polson (2014) and Fulop, Li, and Yu (2015).

We construct monthly bond excess returns on US zero-coupon bonds with maturity 2-, 3-, 4-, and 5-year using the updated dataset of Gurkaynak, Sack, and Wright (2007). The data range from January 1962 to September 2017, in total 669 months. Most studies in bond return predictability focus on predictive regressions for annual excess bond returns at monthly forecasting frequency. Bauer and Hamilton (2017) point out that the bond returns with overlapping holding-period may 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 (2017), we consider one-month holding period and construct non-overlapping monthly excess bond returns.

We take two predictors based on forward rates, i.e., forward spreads (FB) of Fama and Bliss (1987) and the forward factor (CP) proposed by Cochrane and Piazzesi (2005), and construct three predictors based on macroeconomic variables: LN0, LN1, and LN2. LN0 is constructed using the fully revised macroeconomic data by following the approach of Ludvigson and Ng (2009) and is also used by Gargano, Pettenuzzo, and Timmermann (2017); LN1 is constructed using the real-time macroeconomic data by following the approach of Ludvigson and Ng (2009); and LN2 is simply the first principle component of the real-time macroeconomic data. We evaluate statistical out-of-sample predictability using the out-of-sample R-square, R^2_{OS}, of Campbell and Thompson (2008), and evaluate economic out-of-sample predictability using certainty equivalence returns (CERs) by assuming a power-utility investor.

We have some interesting findings. First, we find some statistical evidence of bond return predictability using information contained in forward rates. However, such statistical evidence can hardly be translated into investors' economic gains, no matter how strong is investors' risk aversion. These results are consistent to what Thornton and Valente (2009) and Sarno, Schneider, and Wagner (2016) have found.

Second, when we use the fully revised macroeconomic factor, LN0, we find qualitatively similar results to what Gargano, Pettenuzzo, and Timmermann (2017) have found. That is, significant statistical evidence of bond return predictability can translate into significant investors economic gains. However, as discussed by Ghysels, Horan, and Moench (2017), a real time investor would only have access to real-time macroeconomic data. Therefore, we check whether we can obtain similar significant statistical and economic evidence when LN0 is replaced by either LN1 or LN2. We find that whenever the real-time macro factors are used, both statistical and economic predictability of bond returns vanishes. This result stands in stark contrast to that found by Gargano, Pettenuzzo, and Timmermann (2017). We further show that model combinations do not seem to help to uncover significant statistical and economic evidence of bond return predictability whenever the real-time macro factors are used.

Last, the previous literature has predominantly adopted restrictions on portfolio weights in testing for economic evidence. The usual lower and upper weight bounds for risky bonds are -1 and 2, allowing the possibility of shorting and borrowing. Our previous tests on economic gains also use similar restrictions. However, while such bounds seem natural for equity markets, government bonds are much less risky, resulting in for instance much lower margins in repo transactions backed by these securities. Hence, sophisticated fixed income investors may be able to achieve much more aggressive short and long positions than those implied by these bounds. Therefore, we redo the asset allocation exercise without setting any weight constraints. Interestingly, we find that economic evidence quantitatively greatly improve, especially for the short-maturity bonds. However, it is still not statistically significant when the real-time macro factors are used both in individual predictive models and in model combinations.

Our work makes three main contributions to literature. First, we provide a generic econometric framework allowing real-time Bayesian learning about bond return predictability that takes into account all relevant uncertainties. Thornton and Valente (2012) and Sarno, Schneider, and Wagner (2016) follow classical approaches and therefore ignore such uncertainties. Gargano, Pettenuzzo, and Timmermann (2017) employ Bayesian MCMC methods and do allow for parameter and model uncertainties. However, to investigate out-of-sample predictability, MCMC algorithm needs to be repeatedly run at each time, making it hard to evaluate the speed of convergence and leading to a large computational cost. Our real-time Bayesian learning is tailor-made for sequential inference and is naturally parallel. The only model-dependent requirement of the method is a filtering mechanism for the model in question that provide at least an unbiased likelihood, hence it may be customised to different predictive models easily compared to MCMC methods that are typically more model-dependent.

Second, we call attention to using fully-revised vs. real-time macro information in forecasting exercises. As we face a real-life asset allocation problem, we need to take into account issues related to data revision and publication lag and restrict investors' information set to real-time data only available at each time. The paper by Faust, Rogers, and Wright (2003) is one of the first studies that investigates the difference between fully revised and real-time data. However, they do not find any information advantage from the fully-revised data and argue that this may be due to model misspecification. Ghysels, Horan, and Moench (2017) argue that macroeconomic data revision may result in spurious evidence of bond return predictability. Our results are in line with theirs.

Third, we find interesting result that in bond market it is relatively easy for investors to make extreme investments in the short run with comparison to in equity markets and such extreme investments could improve short-run bond return predictability, though it is still not statistically significant in real time.

The remainder of the paper is organized as follows. Section 2 presents the predictive models and introduces the Bayesian learning approach. Section 3 discusses how we statistically and economically evaluate the predictive performance of each model. Section 4 presents the data and summary statistics. Section 5 provides main empirical results. And Section 6 concludes.

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 is defined as

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

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 m-period later and the yield on a m-period T-bill rate at time t,

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

where m is the holding period in year and $y_t^{(m)}$ is the annualized T-bill rate. In this paper, we assume m is one-month, and n can be 2, 3, 4, or 5 years.

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 parameters. The construction of Equation (3) suggests that $rx_{t+1}^{(n)}$ represents the non-overlapping excess bond return with one-month holding period.

In addition, there is considerable evidence that suggests bond return volatility is time-varying (Gray, 1996; Bekaert, Hodrick, and Marshall, 1997; Bekaert and Hodrick, 2001). Therefore, except the standard model of Equation (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 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 contain information on future 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 with the forecasting horizon. 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. Furthermore, Joslin, Priebsch, and Singleton (2014) provide empirical evidence that macroeconomic variables contain rich information on yields, and Ludvigson and Ng (2009) extract macro factors from a large set of macroeconomic variables and show that these factors have predictive power for future excess bond returns.

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

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

We construct the CP factor following Cochrane and Piazzesi (2005) as follows. At time $t+1$, average excess bond return across maturities is regressed on a set of forward rates at time t,

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

where $\overline{rx}_{t+1} = \frac{1}{4}\sum_{n=2}^{5} rx_{t+1}^{(n)}$ and $\mathbf{f}_t = [f_t^{(1-1/12,1)}, f_t^{(2-1/12,2)}, f_t^{(3-1/12,3)}, f_t^{(4-1/12,4)}, f_t^{(5-1/12,5)}]$. Then the CP factor for time $t+1$ is computed as

$$CP_{t+1} = \hat{\gamma}_0 + \hat{\gamma}\mathbf{f}_{t+1}. \tag{9}$$

Finally, the LN-type macro factor is constructed from a large set of macroeconomic variables using principal component analysis. First, find an optimal combination of principal components and their higher powers, $\hat{\mathbf{F}}_t$, using some statistical criteria, and then build the LN factor as follows

$$LN_{t+1} = \hat{\gamma}_0 + \hat{\gamma}\hat{\mathbf{F}}_{t+1}, \tag{10}$$

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

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

More details on construction of LN-type predictors will be discussed in Section 4.

We only consider single-predictor predictive models and take the predictor to be one of FB, CP or LN. We name each model using the name of its predictor followed by the abbreviation of constant volatility (CV) or stochastic volatility (SV). For example, a model that takes CP as its predictor and assumes stochastic volatility has a name of CP-SV. In model (4), when $\beta = 0$, no predictor is used, and this case is in fact the expectations hypothesis (EH), which will be taken as a benchmark for comparison with the predictive models. We also look into a stochastic volatility version of the EH model, whose performance is even worse than the standard EH model. Therefore, we report results that take EH-CV as the benchmark.

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 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, $y_{1:t} = \{rx_{1:t}^{(n)}, X_{1:t|t}\}$. To account for the fact that CP and LN are updated at each time in the out-of-sample period, we use $X_{1:t_1|t_2}$ to denote the time series of the predictor from time 1 to time t_1 based on information available up to time t_2.

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,

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

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. Investor's belief updating therefore corresponds to updating this joint posterior distribution.

For the linear predictive model of Equation (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 use a particle filter, which is similar to that used in Johannes, Korteweg, and Polson (2014). 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 the 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, $p(rx^{(n)}_{1:t}|\Theta, \mathcal{M}_i)$. Parameter learning can then be implemented as follows, $p(\Theta|y_{1:t}, \mathcal{M}_i) \propto p(y_{1:t}|\Theta, \mathcal{M}_i)p(\Theta|\mathcal{M}_i)$. 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). In contrast to traditional Bayesian MCMC methods such as those used by Gargano, Pettenuzzo, and Timmermann (2017), our Bayesian learning approach does not suffer from convergence issues and 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

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

and an estimate of the marginal likelihood,

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

both of which 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

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

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 investor's 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 suggests

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

Therefore, θ_t is a martingale, indicating that shocks to investor's beliefs on this parameter are not only persistent but also permanent. Thus, in Bayesian learning, the investor 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 investor's initial beliefs or 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 or a truncated normal 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. We find that the model parameters are not sensitive to the selection of the priors. Therefore, we give flat priors to the model parameters. Table 1 provides two sets of the priors that are used in the paper.

2.3. Model Combinations

Model combination is an important tool to handle model uncertainty. Timmermann (2006) argues that model combination can be viewed as a diversification strategy that improves predictive performance in the same manner that asset diversification improves portfolio performance. Rapach, Strauss, and Zhou (2010) and Dangle and Halling (2012) show that model combinations can generate better forecasts than the individual models in forecasting stock returns. In this section, we introduce four model combination schemes for forecasting bond excess returns. Different from MCMC methods, our Bayesian learning algorithm has a natural output of the marginal likelihood of Equation (14), which can be directly used to combine models.

2.3.1. Sequential Best Model (SBM)

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,

$$p_{SBM}(rx_{t+1}^{(n)}|y_{1:t}) = p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i)\mathbb{I}(\max_{\mathcal{M}} p(rx_{1:t}^{(n)}|\mathcal{M}) = \mathcal{M}_i). \tag{17}$$

The best model may change over time, suggesting that a different model may be used for forecasting bond returns at each time.

2.3.2. Bayesian Model Average (BMA)

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,

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

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

$$p(\mathcal{M}_i|rx_{1:t}^{(n)}) = \frac{p(rx_{1:t}^{(n)}|\mathcal{M}_i)p(\mathcal{M}_i)}{\sum_{j=1}^{N} p(rx_{1:t}^{(n)}|\mathcal{M}_j)p(\mathcal{M}_j)}, \tag{19}$$

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

2.3.3. Equal-weighted Model Average (EMA)

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

$$p_{EMA}(rx_{t+1}^{(n)}|y_{1:t}) = \sum_{i=1}^{N} w_{i,t} \times p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i), \tag{20}$$

where N is the number of models considered and $w_{i,t} = 1/N$ for all i and all t.

2.3.4. Utility-based Model Average (UMA)

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 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 $w_{i,t} = p(\mathcal{M}_i|rx_{1:t}^{(n)})$ and is given by

$$p(\mathcal{M}_i|rx_{1:t}^{(n)}) = \frac{CER_{i,t}}{\sum_{j=1}^{N} CER_{j,t}}, \tag{22}$$

in which $CER_{i,t}$ is the certainty equivalent return computed using Equation (31) for the period from the beginning date of out-of-sample to the current time t.

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, t_0, to time t as

$$\widehat{SSE}(t) = \sum_{s=t_0}^{t} (rx_{s+1}^{(n)} - \widehat{rx}_{s+1}^{(n)})^2. \tag{23}$$

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 SSE for expectation hypothesis model is given by

$$\overline{SSE}(t) = \sum_{s=t_0}^{t} (rx_{s+1}^{(n)} - \overline{rx}_{s+1}^{(n)})^2. \tag{24}$$

A natural measure of predictive performance of a model is the out-of-sample R^2 (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 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. Clark and West (2007) show that this statistic performs very well in terms of power and size properties.

3.2. Economic Value and Certainty Equivalence Returns

In evaluating economic predictability, 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},

$$\max_{\omega} \mathbf{E}[U(W_{t+1})|y_{1:t}, \mathcal{M}_i], \tag{26}$$

where $U(\cdot)$ represents the investor's utility function, which is assumed to be a power utility with the relative risk aversion controlled by γ,

$$U(W_{t+1}) \equiv U(\omega_t^{(n)}, rx_{t+1}^{(n)}) = \frac{W_{t+1}^{1-\gamma}}{1-\gamma}, \tag{27}$$

and the portfolio value evolves according to

$$W_{t+1} = (1-\omega_t^{(n)})\exp(r_t^f) + \omega_t^{(n)}\exp(r_t^f + rx_{t+1}^{(n)}), \tag{28}$$

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

Then the expected utility can be computed for each model as follows,

$$\mathbf{E}[U(W_{t+1})|y_{1:t}, \mathcal{M}_i] = \int U(\omega_t^{(n)}, rx_{t+1}^{(n)}) p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i) drx_{t+1}^{(n)}, \tag{29}$$

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 Equation (29), 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

$$\hat{w}_t^{(n)} = \arg\max \frac{1}{M} \sum_{j=1}^{M} \left\{ \frac{\left[(1-\omega_t^{(n)})\exp(r_t^f) + \omega_t^{(n)}\exp(r_t^f + rx_{t+1}^{(n),(j)})\right]^{1-\gamma}}{1-\gamma} \right\}. \tag{30}$$

The above portfolio weight in Equation (30) 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 EH benchmark 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

$$CER = \left(\frac{\sum_{t=1}^{T} \hat{U}_t}{\sum_{t=1}^{T} \bar{U}_t} \right)^{\frac{1}{1-\gamma}} - 1. \tag{31}$$

4. Data and Summary Statistics

We construct monthly yields on US zero-coupon bonds with maturity 2-, 3-, 4-, and 5-year using the updated yield dataset of Gurkaynak, Sack, and Wright (2007). Most studies in bond return predictability focus on predictive regressions for annual excess bond returns in monthly forecasting frequency, that is, m is one year 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 (2017), we consider one-month holding period and construct non-overlapping monthly excess bond returns. This implies that m is equal to one month in Equations (2) and (3).

Ghysels, Horan, and Moench (2017) argue that macroeconomic data revision may introduce bond return predictability. Therefore, we use two types of macroeconomic data, one is fully-revised, and the other is real-time, to construct LN-type macro predictors. LN is a linear combination of principal components (PCs) and their second and third powers. We construct our first LN-type macro predictor, LN0, using the fully-revised macroeconomic data, downloaded from St. Louis Fed, by relying on the optimal subset of PCs and their powers recommended Ludvigson and Ng (2009).[1]

We also construct two LN-type macro predictors, LN1 and LN2, using the real-time macroeconomic data published by McCracken and Ng (2016).[2] Due to publication delay,

[1] This subset is $[\hat{F}_{1,t}, \hat{F}_{1,t}^3, \hat{F}_{3,t}, \hat{F}_{4,t}, \hat{F}_{8,t}]$. The fully-revised macro data are downloaded at https://research.stlouisfed.org/econ/mccracken/fred-databases.

[2] The vintage data can be downloaded at https://research.stlouisfed.org/econ/mccracken/fred-

at each month t we can only have the observation of each macroeconomic variable for month $t-1$. We denote this most recent observation of variable i as $Macro^i_{t-1|t}$, $i \in \{1:I\}$. We then construct LN1 as follows. At each month $t+1$ in real time, we can observe $rx_{\{1:t+1\}}$ and the macroeconomic panel data, $Macro^{1:I}_{\{0:t\}|t+1}$. We first determine the number of principal components and extract them from the real-time macro panel using the method proposed by Bai and Ng (2002). At the beginning of the out-of-sample period, we then pin down the optimal subset from the first three powers of all PCs using Bayesian Information Criterion, resulting in $\hat{\mathbf{F}}_{t+1} = [\hat{F}_{3,t+1}, \hat{F}^3_{6,t+1}]$. We finally re-build LN1 at each following time. LN2 is constructed simply using the first PC as discussed by Ghysels, Horan, and Moench (2017).

Our sample spans from January, 1962 to September, 2017. In total, there are 669 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 1.32% with a standard deviation of 2.82% for the 2-year bond, whereas it increases to 2.11% with a standard deviation of 5.98% 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 0.55 and 16.4, respectively, whereas they are only 0.02 and 6.96 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 predictors: FB, CP, and LNs. We find that (1) the 2-year FB is positively skewed, whereas other FBs are negatively skewed,

databases.

and kurtosis of FBs decrease with respect to maturity. The FB factors are persistent, as the first-order autocorrelation ranges from 0.88 (2-year) to 0.92 (5-year); (2) the CP factor is positively skewed and leptokurtic, and its autocorrelation is about 0.70; (3) the three LN factors are very different: the kurtosis of real-time LNs is much larger than that of LN0, and its skewness display different sign and magnitude. Panel C reports correlations between full-sample predictors. The FB factors are highly correlated with each other, and weakly correlate with the three types of macro factors; The CP factor has moderate correlations with the FB factors and weak correlation with the three types of macro factors. The correlations among the three macro factors are weak in general.

It is important to note that when we empirically implement Bayesian learning in the out-of-sample period, both CP and LNs are reconstructed at each time using information available only up to that time in order to avoid any hindsight problems.

5. Empirical Results

5.1. Parameter Learning and Sequential Model Comparison

Different from batch estimation methods such as Bayesian MCMC methods, our Bayesian learning approach provides us with the whole picture of how parameter posteriors evolve over time with respect to the accumulation of information for each model. In this section, we focus on a stochastic volatility model and a constant volatility model, both of which take the FB factor as their only predictor (i.e., FB-SV and FB-CV). Figure 2 presents the sequential learning results of the fixed parameters for FB-SV, and Figure 3 presents the sequential learning of the fixed parameters for FB-CV, on 3-year excess bond returns. For each parameter, the posterior mean (solid line) and the 5th and 95th percentile credible interval (dashed lines) are plotted.

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, indicating decreasing of parameter uncertainties.

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 FB-SV and FB-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. 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 5th and 95th percentile credible interval to sufficiently narrow down.

Third, the last panel of Figure 2 presents the sequential estimates of stochastic volatility for FB-SV model. 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 5th and 95th percentile credible intervals. There is a large spike of volatility around the beginning of 1980s.

Fourth, the learning process of σ_{rx} in FB-CV in Figure 3 reveals potential 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. This volatility estimate could be seen as a much smoothed version of the stochastic volatility in Figure 2.

Last, thanks to its recursive nature, our Bayesian learning approach produces the sequential marginal likelihood at each time for each model 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 3 presents the sequential log Bayes factors between FB-SV and FB-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, FB-SV performs nearly the same as FB-CV. Second, as the market information accumulates over time, the data begin to strongly favour the stochastic volatility model. Third, most of the up-moves in Bayes factors happen during market turmoils. This phenomenon is particularly striking around 1980 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, the stochastic volatility model performs better than the constant volatility model on the 5-year excess bond returns before 1980, whereas afterwards, its performance becomes much stronger on the 2-year excess bond returns.

5.2. Evidence of Bond Return Predictability

Due to availability of macroeconomic vintage data, we set our out-of-sample period from August 1999 to September 2017, in total, 218 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} for evaluating its statistical predictive performance. In investigating economic evidence, we first restrict the portfolio weight, $\omega_t^{(n)}$, in between -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). We then relax this restriction in Subsection 5.2.4. For most part of this section, we use the first set of priors in Table 1 in our Bayesian learning. We will have a robustness check on sensitivity to priors in the last subsection.

5.2.1. Information Contained in Forward Rates

Fama and Bliss (1987) find that FB can predict future excess bond returns and that its forecasting power increases with the forecasting horizon, and Cochrane and Piazzesi (2005) show that the whole term structure of forward rates, CP, can capture more than 30% of the variation of excess bond returns over the period from January 1964 to December 2003. However, these results are pure in-sample statistical evidence. In this paper, we implement an out-of-sample investigation and see whether FB and CP have any predictive power to excess bond returns both statistically and economically.

Panel A of Table 3 presents R^2_{OS}'s from the models using FB and CP as predictors. We have the following findings. First, R^2_{OS}s based on FB (FB-CV and FB-SV) increase with respect maturity and only those R^2_{OS}s for long-maturity bond returns are statistically significant; Second, it seems that the introduction of stochastic volatility is unhelpful whenever FB is used. For example, for the 4-year excess bond returns, the R^2_{OS} from FB-CV is 2.45% and statistically significant, whereas the corresponding R^2_{OS} from FB-SV is only 1.86% and marginally significant, and for 5-year excess bond returns. both FB-CV and FB-SV generates similar R^2_{OS} and statistically significant, 2.72% vs. 2.81%. Third, the results from the models based on CP are different because CP-SV performs better than CP-CV and those R^2_{OS} for short-maturity excess bond returns are (marginally) significant.

The results from Panel A point to some statistical evidence of out-of-sample bond return predictability using information contained in forward rates. We then investigate whether such statistical evidence can translate into investors' economic gains. Our investor is Bayesian, who takes into account all relevant uncertainty when maximizing her expected utility in Equation (26). We compute the corresponding certainty equivalent returns (CERs) for each model using formula (31), and test if the annualized CER values are statistically greater than zero using a one-sided Diebold-Mariano test as discussed in Gargano, Pettenuzzo, and Timmermamm (2017).

Panels B, C, and D present CERs from the models using FB and CP as predictors by setting the coefficient of relative risk aversion at 3, 5, and 10, respectively. We find that no matter which model is used and which maturity is considered, all CERs are not statistically larger than zero. However, we do find that whenever investors become more risk-averse, most CERs for medium- and/or long-maturity bond returns are positive. These results indicate that any statistical evidence based on forward rates is hard to translate into economic gains, consistent to what Thornton and Valente (2012), Sarno, Schneider, and Wagner (2016), and Gargano, Pettenuzzo, and Timmermamm (2017) have found.

5.2.2. *Revised and Real-Time Macro Information*

Several studies show that macroeconomic variables contain rich information on future excess bond returns beyond information in the yield curve (Ludvigson and Ng, 2009; Cooper and Priestly, 2009; Huang and Shi, 2014; Joslin, Priebsch, and Singleton, 2014; Jiang and Tong, 2017). Gargano, Pettenuzzo, and Timmermamm (2017) implement an out-of-sample test and show that statistical evidence based on macroeconomic information can translate into investors' economic gains. However, most of these works use the fully-revised macroeconomic variables and ignore data revision and publication delay. A recent paper by Ghysels, Horan, and Moench (2017) argue that macroeconomic data revision may introduce spurious bond return predictability. To check whether macroeconomic variables contain information on future excess bond returns and data revision makes a difference, we construct three LN-type macro factors as discussed in Section 4, LN0 from fully-revised macroeconomic data, and LN1/LN2 from real-time macroeconomic data.

Panel A of Table 4 presents R^2_{OS}s from the models using LN-type factors as predictors, and Panels B, C, and D report CERs from these models with the coefficient of relative risk aversion equal to 3, 5, and 10, respectively. We find strong statistical evidence of bond return predictability when combining the fully revised macro factor, LN0, and stochastic volatility. For example, the R^2_{OS}s from LN0-SV are 3.08%, 3.74%, 3.83%, and 3.62% for 2-, 3-, 4-, and 5-year bond returns, respectively, and are all statistically significant. And importantly, this statistical predictability can translate into significant economic gains for 4- and 5-year excess bond return when the coefficient of risk aversion is set at 5 (Panel C), and for 2-, 3-, and 4-year excess bond return when the coefficient of risk aversion is set at 10 (Panel D). These results are qualitatively similar to those found by Gargano, Pettenuzzo, and Timmermamm (2017). In fact, we also find similar results from using LN0 by setting the out-of-sample period ranging from January 1990 to December 2015, which is the same as that in Gargano, Pettenuzzo, and Timmermamm (2017).

However, we have to take into account the issue of macro data revision and publication delay issue in analysis of bond return predictability. For this reason, we implement the same empirical investigation using two real-time LN-type macro factors, LN1 and LN2, instead of LN0. We find strikingly different results. No matter which real-time macro factors we use, whether we introduce stochastic volatility, or how strong the investor's risk aversion is, there exist neither statistical evidence nor economic evidence of bond return predictability. It seems that there is evidence that the first principle component extracted from a large panel of macroeconomic data (LN2) works better than that constructed using Ludvigson and Ng (2009) approach (LN1).

5.2.3. Model Combinations

We now move on to check whether model combinations can help find any statistical and/or economic evidence of bond return predictability. Given that the three LN-type predictors are constructed completely differently, we implement our model combination schemes for three groups of models, each of which contains only one macro predictor. Panel A of Table 5 presents the model combination results from using predictors of FB, CP, and LN0. We find that the R^2_{OS}s from BMA, EMA and UMA are in general larger than those in Table 3 and Table 4 and are significant nearly for all-maturity excess bond returns. Furthermore, we find that the CERs from EMA and UMA are positive and significant for 4- and 5-year excess bond returns when the coefficient of relative risk aversion sets to 5, and he CERs from BMA, EMA, and UMA are positive and significant for 2-, 3- and 4-year excess bond returns when the coefficient of relative risk aversion sets to 10.

Panels B and C presents the model combination results from using predictors of FB and CP, together with LN1 and LN2, respectively. We find that whenever the real-time macro factors are used, the statistical and economic evidence we have found in Panel A vanishes, though we notice a little statistical evidence in Panel C from SBM and UMA for 2-year bond returns when LN2 is used.

5.2.4. Extreme Investments and Predictive Performance

Up to now, we have restricted portfolio weight in between -1 and 2. However, different from the equity market, it may be feasible for investors to extreme positions in the bond market, facilitated by Repo agreements for instance. Therefore, we allow investors to take their investment decision without any restrictions on portfolio weight. In the out-of-sample period, the maximum weights are 97, 38, 23, and 17, and the minimum weights are -53, -18, -10, and -7, on 2-, 3-, 4-, and 5-year bonds, respectively, when the coefficient of the relative risk aversion is set to 5. The average weights on these bonds are 8.12, 3.56, 2.21, and 1.61, respectively, decreasing with respect to maturity.

Table 6 presents CERs from all individual models with the coefficient of risk aversion at 5. We find that with comparison to those in Table 3 and Table 4, nearly all CERs greatly improve, especially those for short-maturity bonds (2-year and 3-year). The average annualized CERs are 2.99%, 1.52%, 0.80%, and 0.78% for 2-, 3-, 4-, and 5-year bonds, respectively. This suggest that without imposing any restrictions on portfolio weight, the economic evidence becomes more pronounced, especially for the short-maturity risky bonds. The investment on the 2-year bond is now the most profitable across all maturities. It seems a new result in the literature on bond return predictability and contrast with the general finding on equities where constraints on the portfolio weights tend to improve performance out of sample (see, e.g., Pettenuzzo, Timmermann, and Valkanov, 2014). The reason may be that government bonds are much less risky compared to equity, hence even smaller fluctuations in the conditional expected return and/or conditional volatility suggest wildly varying portfolio weights.

However, we still find that CERs from FB- and CP-based models are hardly statistically significant, and even though the CERs of 2-year bond from LN0-based models are statistically significant, they become statistically insignificant whenever the revised macro factor (LN0) was replaced by the real-time macro factors (LN1 and LN2), with one exception of LN2-CV for 2-year bond. Similar results can be found when we set the coefficient of risk aversion equal to 3 and 10.

Table 7 presents CERs from model combinations with the coefficient of risk aversion at 5. As before, we consider the same three groups of models. We find very similar results. With comparison to Table 5 and Table 6, CERs for each maturity improve greatly, especially for 2-year bond. When models based on LN0 are grouped together with models based on FB and CP, we find that BMA, EMA, and UMA generate high and statistically significant CERs for 2-year bond, and EMA and UMA generate statistically significant CERs for 3-year bond. However, consistent to what we have found in Table 5, such significance vanishes when the real-time macro factor, LN1 or LN2, is grouped together with FB and CP. We find qualitatively similar results when the coefficient of risk aversion is equal to 3 or 10.

5.2.5. 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 obtain nearly the same results as those in the previous subsections. Therefore, we conclude that our results are not sensitive to the choice of priors at all.

6. Concluding Remarks

The paper studies both statistical and economic evidence of out-of-sample bond return predictability for a real-time Bayesian investor who learns about parameters, hidden states, and predictive models over time when new information becomes available. We take two predictors based on forward rates, i.e., forward spreads (FB) of Fama and Bliss (1987) and the forward factor (CP) proposed by Cochrane and Piazzesi (2005), and construct three predictors based on macroeconomic variables: one is constructed using the fully revised macroeconomic data by following the approach of Ludvigson and Ng (2009) and is also used by Gargano, Pettenuzzo, and Timmermann (2017), and the other two are constructed using the real-time macroeconomic data. We compare the predictive performance of expectations hypothesis and predictive models, and evaluate statistical out-of-sample predictability using the out-of-sample R-square, R^2_{OS}, of Campbell and Thompson (2008), and evaluate economic out-of-sample predictability using certainty equivalence returns (CERs) by assuming a power-utility investor.

Most studies in bond return predictability focus on predictive regressions for annual excess bond returns with monthly forecasting horizon. Such overlapping returns introduce 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 (2017), we consider one-month holding period and construct non-overlapping monthly excess bond returns. Using both yield data and revised and real-time macroeconomic data ranging from January 1962 to September 2017, we find some statistical evidence using information contained in forward rates. However, such statistical predictability can not generate any economic value for investors. Furthermore, statistical and economic evidence from fully revised macroeconomic data completely vanishes when real-time macroeconomic information is used. We also show that extreme investments in bonds could improve short-run bond return predictability, though such predictability is not significant in real time.

REFERENCES

Bai, J., and S. Ng., 2002. "Determining the Number of Factors in Approximate Factor Models.". *Econometrica* 70, 135-172.

Bauer, M. and Hamilton, J., 2017. "Robust bond risk premia". *Review of financial studies*, forthcoming.

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.

Faust, J., Rogers, J., and Wright, J., 2003. "Exchange rate forecasting: the errors we've really made". *Journal of International Economics* 60, 35–69.

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". *Management Science*, forthcoming.

Ghysels, E., Horan, C., and Moench, E., 2017. "Forecasting through the Reaview Mirror: Data revisions and bond return predictability", *Review of Financial Studies*, forthcoming.

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.*

Jiang, F. and Tong, G., 2017. “Monetary policy uncertainty and bond risk premium”. *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.

Lin, H., Liu, W., Wu, C., and Zhou, G., 2017. “Prediction of treasury bond returns”, working paper.

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

McCracken, M. and Ng, S., 2016, “FRED-MD: A monthly database for macroeconomic research”. *Journal of Business and Economic Studies* 34, 574–589.

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

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.

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 presents two sets of prior distributions to be used in the paper. The linear model is given in Equation (4). Parameters for the linear models are: α, β, and σ_{rx} The stochastic volatility model is given in Equations (5) and (6). Parameters for the SV models are: α, β, μ, ϕ, and σ_h.

Table 2: **Full-sample Summary Statistics**

Panel A: Execess Bond Returns

	2-Year	3-Year	4-Year	5-Year
Mean	1.321	1.642	1.902	2.114
St.dev	2.822	3.960	4.992	5.979
Skew	0.550	0.234	0.071	0.023
Kurt	16.384	11.588	8.468	6.959
AC(1)	0.167	0.149	0.132	0.116

Panel B: Predictors

	FB2	FB3	FB4	FB5	CP	LN0	LN1	LN2
Mean	0.104	0.128	0.147	0.162	0.146	0.146	0.146	0.146
St.dev	0.095	0.110	0.122	0.132	0.197	0.265	0.107	0.092
Skew	0.015	-0.233	-0.265	-0.212	0.668	0.402	-1.372	1.254
Kurt	4.016	3.642	3.283	2.965	4.466	3.469	9.876	6.639
AC(1)	0.880	0.899	0.913	0.923	0.703	0.471	0.542	0.718

Panel C: Correlations

	FB2	FB3	FB4	FB5	CP	LN0	LN1	LN2
FB2	1.000	0.969	0.914	0.859	0.490	-0.151	0.307	0.052
FB3		1.000	0.985	0.955	0.500	-0.118	0.404	0.116
FB4			1.000	0.992	0.510	-0.082	0.463	0.169
FB5				1.000	0.512	-0.054	0.498	0.207
CP					1.000	0.099	0.108	0.212
LN0						1.000	-0.048	0.465
LN1							1.000	-0.020
LN2								1.000

This table presents the summary statistics of bond excess returns and full-sample predictors. Panel A reports 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 the predictors. Panel C shows the correlation matrix of the predictors. Full-sample data is from January 1962 to September 2017.

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

	A. R^2_{OS}				C. CER: $\gamma = 5$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
FB-CV	-0.17	1.42*	2.45**	2.72**	-0.96	-0.80	0.05	0.95
CP-CV	0.39*	0.59*	0.48	1.42	-0.68	-0.84	-0.59	0.00
FB-SV	-2.17	0.25	1.86*	2.81**	-1.03	-0.89	-0.04	0.87
CP-SV	1.97**	1.08*	1.70	2.15*	-0.06	-0.28	0.75	1.40
	B. CER: $\gamma = 3$				D. CER: $\gamma = 10$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
FB-CV	-0.97	-1.24	-0.96	-0.22	-0.31	0.48	1.08	1.22
CP-CV	-0.69	-1.15	-1.74	-1.14	-0.12	0.11	0.07	0.24
FB-SV	-1.07	-1.32	-1.17	-0.36	-0.35	0.35	0.70	0.87
CP-SV	-0.01	-0.61	-0.52	0.48	0.40	0.74	1.17	1.10

This table presents the out-of-sample R-squares, R^2_{OS}, and annualized CERs for the predictive models based on FB and CP. The portfolio weight is restricted in between -1 and 2. The statistical significance of R^2_{OS} is measured using the Clark and West (2007) statistic, and the statistical significance of CERs is measured by the one-sided Diebold-Mariano statistic. * denotes significance at 10% level, ** significance at 5% level, and *** significance at 1% level. The out-of-sample period is from August 1999 to September 2017.

Table 4: **Out-of-Sample Predictability: Macro Factors**

	A. R^2_{OS}				C. CER: $\gamma = 5$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
LN0-CV	-13.40	-4.44	-0.61	0.64*	0.24	0.91	2.19***	2.90***
LN0-SV	3.08**	3.74**	3.83**	3.62**	0.24	0.89	2.01**	2.55**
LN1-CV	-38.04	-25.25	-17.35	-12.03	-0.38	-0.47	0.06	0.17
LN2-CV	0.13	-0.44	-0.11	0.32	-0.03	0.09	0.33	0.71
LN1-SV	-15.11	-22.45	-19.05	-10.28	-0.66	-2.04	-1.94	-1.22
LN2-SV	-0.28	-1.32	-1.60	-1.20	-0.13	-0.16	0.00	-0.17
	B. CER: $\gamma = 3$				D. CER: $\gamma = 10$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
LN0-CV	0.20	0.42	0.82	1.54	0.92**	2.02***	2.40***	2.15*
LN0-SV	0.15	0.53	0.68	1.43	0.89**	1.97***	2.29**	1.93*
LN1-CV	-0.25	-0.63	-0.72	-0.33	-0.01	-0.05	-0.03	-0.01
LN2-CV	0.00	-0.08	0.06	0.63	0.47**	0.31	0.17	0.20
LN1-SV	-0.51	-2.18	-3.26	-2.71	-0.45	-0.70	-0.75	-0.45
LN2-SV	-0.04	-0.37	-0.97	-0.67	0.40	0.77	0.14	-0.24

This table presents the out-of-sample R-squares, R^2_{OS}, and annualized CERs for the predictive models based macro factors, LN0, LN1, and LN2. The portfolio weight is restricted in between -1 and 2. The statistical significance of R^2_{OS} is measured using the Clark and West (2007) statistic, and the statistical significance of CERs is measured by the one-sided Diebold-Mariano statistic. * denotes significance at 10% level, ** significance at 5% level, and *** significance at 1% level. The out-of-sample period is from August 1999 to September 2017.

Table 5: **Out-of-Sample Predictability: Model Combinations**

Panel A: FB, CP, and LN0

	R^2_{OS}				CER: $\gamma = 5$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	0.61**	-0.78	0.47	3.36*	-0.32	-0.18	0.39	1.61
BMA	3.71**	3.83**	3.33*	2.46	0.34	0.86	1.49**	1.00
EMA	5.79**	5.24**	4.97**	4.78**	-0.01	0.39	1.39**	2.05**
UMA	3.33***	0.58**	2.99**	3.23*	0.30	0.92	2.05**	2.51**

	CER: $\gamma = 3$				CER: $\gamma = 10$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	-0.33	-0.53	-0.95	0.51	0.29	0.83	0.69	0.78
BMA	0.25	0.42	0.40	0.05	0.99***	1.99***	1.95**	0.73
EMA	0.00	0.11	0.18	0.86	0.53**	1.43***	1.57**	1.49*
UMA	0.31	0.50	0.72	1.27	0.91***	2.04***	2.30***	1.79*

Panel B: FB, CP, and LN1

	R^2_{OS}				CER: $\gamma = 5$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	-2.57	-25.48	-21.90	-3.80	-0.57	-1.46	-1.84	-1.73
BMA	-0.49	0.22	1.77*	2.32*	-0.78	-0.90	0.15	0.59
EMA	-0.47	-0.40	1.69*	2.80*	-0.40	-0.52	0.16	0.69
UMA	1.08*	-1.33	1.15	0.40	-0.42	-0.75	-0.06	-0.35

	CER: $\gamma = 3$				CER: $\gamma = 10$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	-0.56	-1.56	-2.78	-2.20	-0.25	-0.89	-1.59	-1.82
BMA	-0.79	-1.33	-0.86	-0.68	-0.23	0.33	0.62	0.59
EMA	-0.38	-0.65	-0.61	-0.08	-0.08	0.35	0.67	0.72
UMA	-0.43	-0.77	-1.17	-1.20	0.04	0.08	0.58	-0.06

Panel C: FB, CP, and LN2

	R^2_{OS}				CER: $\gamma = 5$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	1.72**	-0.16	0.61	1.21	0.04	-0.48	-0.04	0.69
BMA	-0.41	0.26	1.86*	2.46*	-0.39	-0.91	0.33	0.69
EMA	1.98*	1.81*	1.94*	2.15*	-0.22	-0.17	0.48	0.84
UMA	2.01**	-1.82	1.66	0.08	-0.34	-0.57	0.30	-0.34

	CER: $\gamma = 3$				CER: $\gamma = 10$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	0.05	-0.71	-1.31	-0.33	0.43	0.36	0.32	0.18
BMA	-0.36	-1.31	-0.66	-0.49	0.01	0.40	0.74	0.74
EMA	-0.22	-0.33	-0.37	0.04	0.23	0.64*	0.71	0.71
UMA	-0.38	-0.79	-0.91	-0.90	0.22	-0.03	0.65	-0.15

This table presents the out-of-sample R-squares, R^2_{OS}, and annualized CERs for the four model combination schemes. The portfolio weight is restricted in between -1 and 2. The statistical significance of R^2_{OS} is measured using the Clark and West (2007) statistic, and the statistical significance of CERs is measured by the one-sided Diebold-Mariano statistic. * denotes significance at 10% level, ** significance at 5% level, and *** significance at 1% level. The out-of-sample period is from August 1999 to September 2017.

Table 6: **Out-of-Sample Predictability: Extreme Investments**

	2-year	3-year	4-year	5-year
FB-CV	-0.09	1.12	2.17	2.47
CP-CV	1.05	0.70	0.42	0.81
LN0-CV	7.03**	4.90	2.80	1.19
LN1-CV	0.68	0.45	0.37	0.33
LN2-CV	1.51**	0.58	0.34	0.41
FB-SV	0.28	0.75	0.85	1.21
CP-SV	2.26	-0.14	-0.04	0.54
LN0-SV	11.03**	6.67	2.87	2.00
LN1-SV	1.95	-0.17	-0.77	-0.22
LN2-SV	4.20	0.33	-0.99	-0.92

This table presents annualized CERs for all individual predictive models without any restrictions on portfolio weight. The coefficient of the relative risk aversion is set to 5. The statistical significance of CERs is measured by the one-sided Diebold-Mariano statistic. * denotes significance at 10% level, ** significance at 5% level, and *** significance at 1% level. The out-of-sample period is from August 1999 to September 2017.

Table 7: **Out-of-Sample Predictability: Model Combinations with Unbounded Weights**

	2-year	3-year	4-year	5-year
Panel A: FB, CP, and LN0				
SBM	2.01	1.56	-2.65	-0.71
BMA	12.64**	7.41	1.78	0.17
EMA	4.02***	3.72**	3.43*	3.01
UMA	8.31***	5.95*	2.58	2.62
Panel B: FB, CP, and LN1				
SBM	-3.98	-4.80	-5.57	-5.64
BMA	1.62	0.20	0.08	0.45
EMA	1.02	1.33	1.69	1.77
UMA	1.05	-0.83	0.47	-0.65
Panel C: FB, CP, and LN2				
SBM	0.08	-1.37	-1.77	-0.50
BMA	3.34	0.72	-0.09	0.70
EMA	1.75*	1.72	1.44	1.57
UMA	2.14	-0.09	1.20	-0.22

This table presents annualized CERs for the four model combination schemes without any restrictions on portfolio weight. The coefficient of the relative risk aversion is set to 5. The statistical significance of CERs is measured by the one-sided Diebold-Mariano statistic. * denotes significance at 10% level, ** significance at 5% level, and *** significance at 1% level. The out-of-sample period is from August 1999 to September 2017.

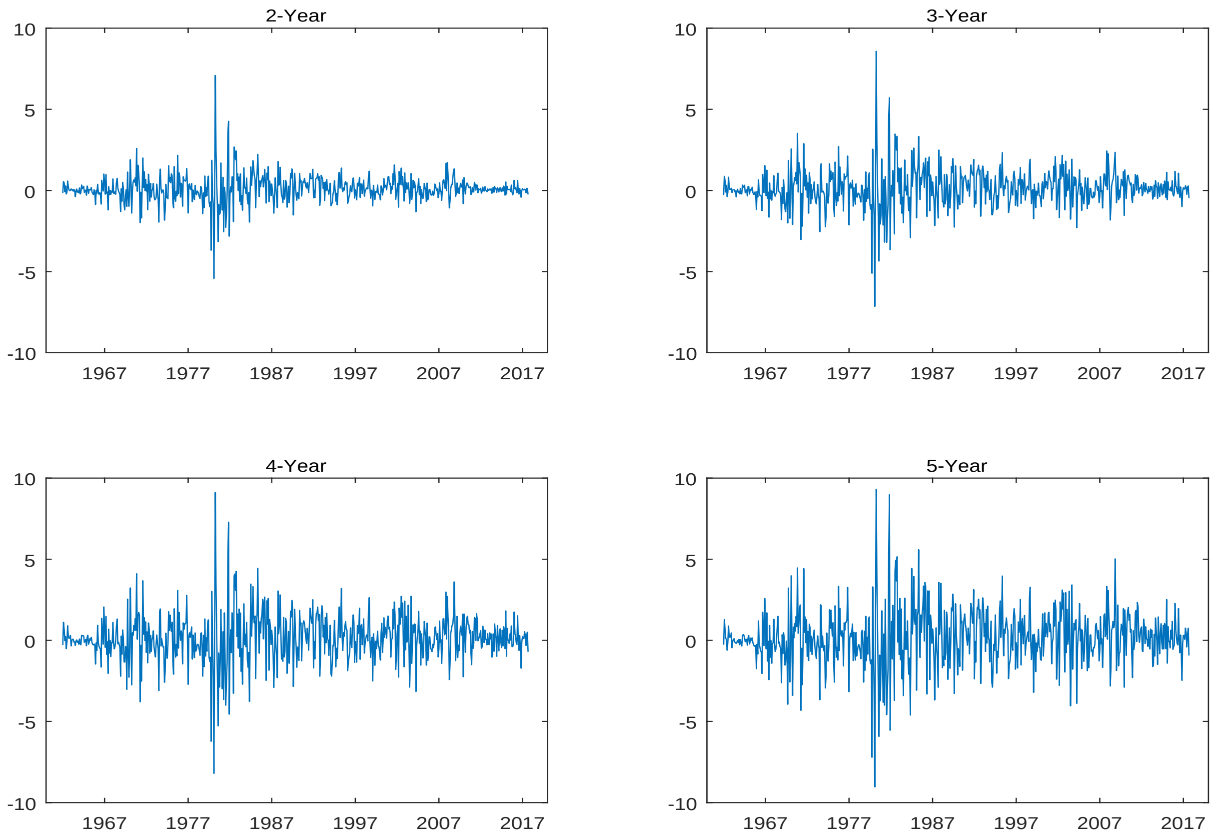

Figure 1: **The Time Series of Excess Bond Returns**
This figure plots the time series of 4 excess bond returns (in percentage), from Jan, 1962 to Sep, 2017.

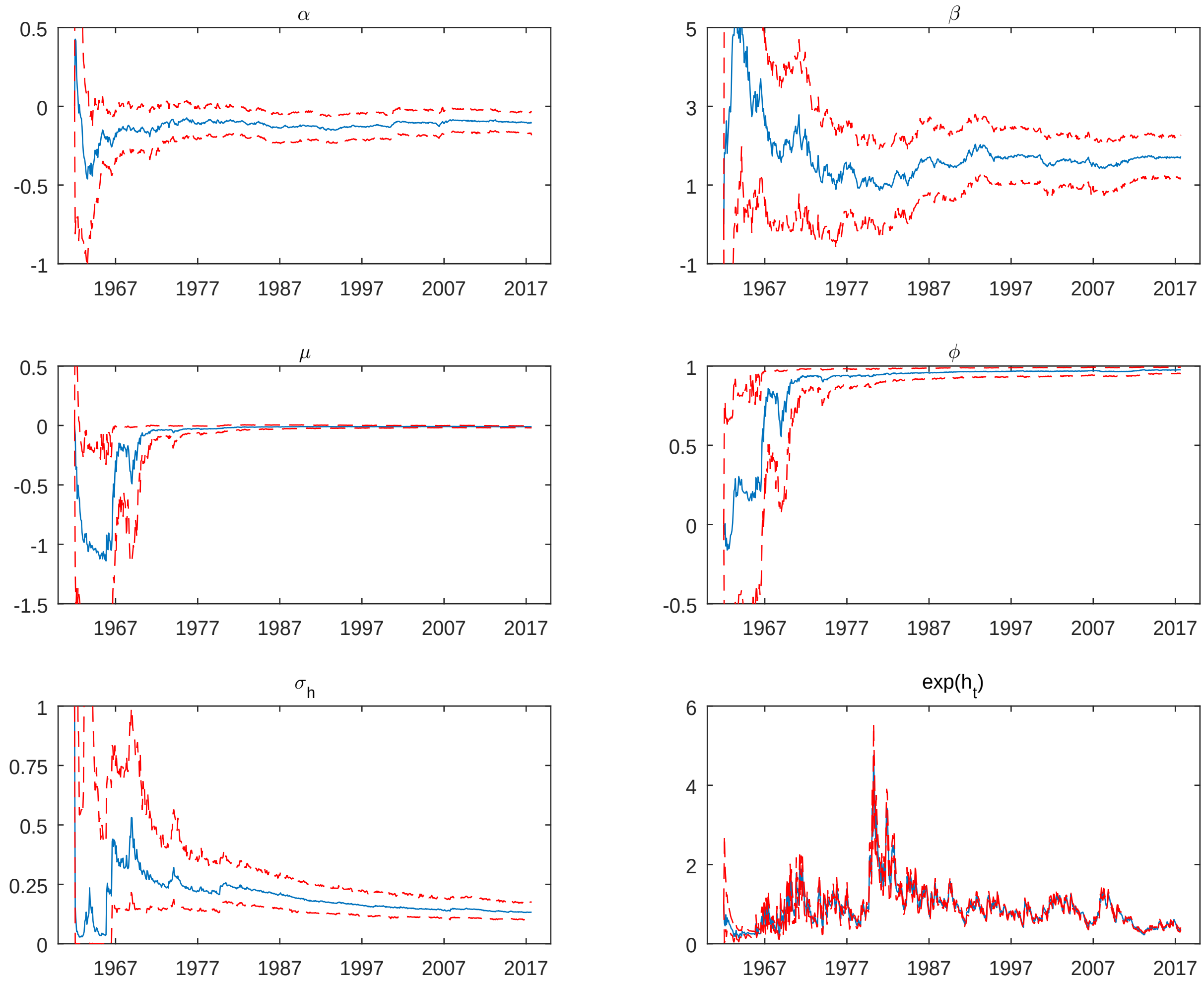

Figure 2: **Parameter Learning for FB-SV**

The figure shows time series parameter estimates of stochastic volatility model with FB predictor for 3-year bond excess returns. 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. Sample is from Jan, 1962 to Sep, 2017.

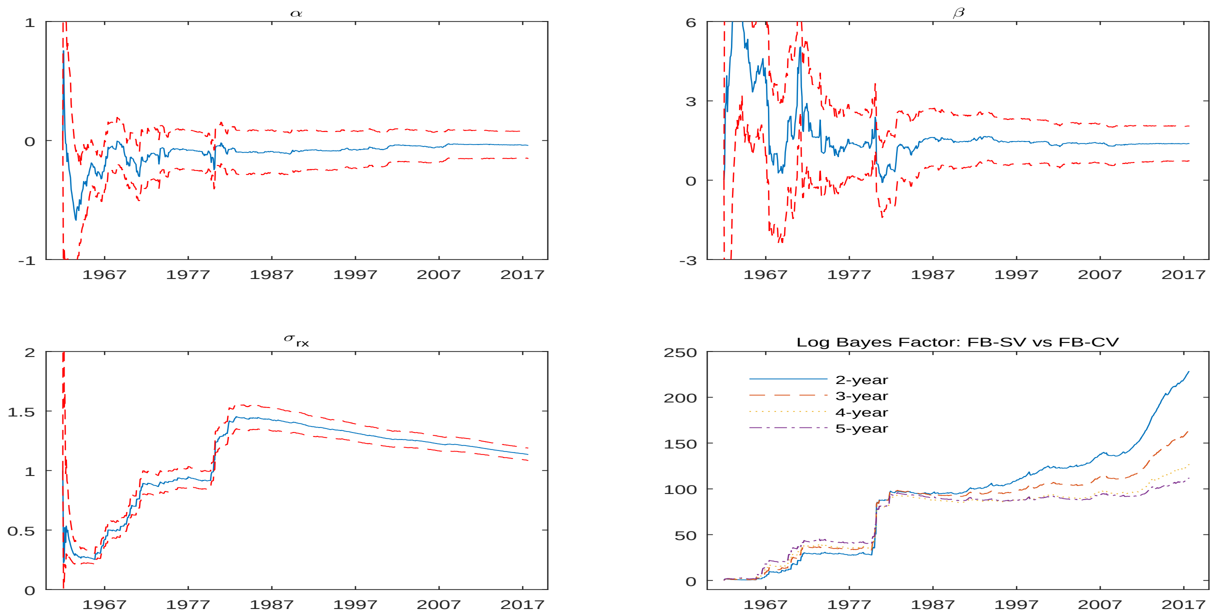

Figure 3: **Parameter Learning for FB-CV and Bayes Factor**

The figure shows time series parameter estimates of constant volatility model with FB predictor for 3-year bond excess returns. The linear model form is given in equation (4). The two dashed lines are 5-th and 95-th percentiles of estimate distribution. The solid line is the mean estimate for each parameter. Last panel shows the bayes factor for all 4 maturities. Sample is from Jan, 1962 to Sep, 2017.