THREE ESSAYS ON EMPIRICAL ASSET PRICING

A dissertation submitted

in partial fulfillment of the requirements for the degree of

PHD IN BUSINESS ADMINISTRATION

and for the degree of

DOCTEUR EN SCIENCES DE GESTION

DE L'ECOLE DOCTORALE

« ECONOMIE, MANAGEMENT, MATHEMATIQUES, PHYSIQUE ET SCIENCES INFORMATIQUES »

ED 405

UNIVERSITE PARIS-SEINE

Defended publicly the 25th of June 2019 by

Runqing WAN

JURY

Andras FULOP	Co-Supervisor	Associate Professor, ESSEC Business School (Cergy, France)
Junye LI	Co-Supervisor	Professor, ESSEC Business School (Singapore, Singapore)
Roméo TEDONGAP	Chair	Professor, ESSEC Business School (Cergy, France)
Serge DAROLLES	Referee	Professor, Université Paris-Dauphine (Paris, France)
Jun YU	Referee	Professor, Singapore Management University (Singapore, Singapore)
Elise GOURIER	Examiner	Assistant Professor, ESSEC Business School (Cergy, France)
Jean-Paul RENNE	Examiner	Assistant Professor, Université de Lausanne (Lausanne, Switzerland)
Guillaume ROUSSELLET	Examiner	Assistant Professor, McGill University (Montreal, Canada)

Acknowledgements

I owe a huge debt of gratitude to my supervisors, Professor Andras Fulop and Professor Junye Li, for their patience, motivation, and great knowledge. I truly appreciate all their contributions of time, ideas, and funding. I feel so honored and blessed to be their student.

I am very grateful to Professors Serge Darolles, Elise Gourier, Jean-Paul Renne, Guillaume Roussellet, Roméo Tédongap, and Jun Yu for accepting to be in my dissertation committee. I deeply thank Professors Sridhar Arcot, Laurent Bach, Romain Boulland, Jieying Hong, Laurence Lescourret, Jocelyn Martel, José-Miguel Gaspar, Patrice Poncet, Sofia Ramos, Peng Xu, and Vijay Yadav in ESSEC Finance Department. I also thank Lina Prevost and Christine Gil from the Ph.D. Program Office.

I would like to express my appreciation to all my friends and colleagues. There are so many of them, and I learn so much from each person. May our friendship last forever.

Lastly, and most importantly, I thank my parents for their love, trust, and support. I dedicate this thesis to them.

Abstract

This doctoral thesis includes three essays investigating several topics in empirical asset pricing. Essay 1 examines statistical and economic evidence of out-of-sample Treasury bond return predictability for a real-time Bayesian investor who learns about parameters, states, and predictive models over time. We can identify some statistical evidence using information in forward rates, however such statistical predictability can not generate any economic value for investors. Furthermore, the strong statistical and economic evidence from using fully revised macroeconomic data vanishes when real-time and survey-based macroeconomic information is used. We also show that highly levered investments can improve short-run bond return predictability.

Essay 2 investigates bond risk premia in the framework of predictive systems. Different from the traditional linear predictive models, predictive systems allow predictors to be imperfectly correlated with conditional expected returns, and could incorporate prior beliefs on the negative correlation between unexpected and expected returns. We find that predictive systems can deliver stronger evidence of pre-

dictability than linear predictive models. Furthermore, bond risk premia inferred by predictive systems are countercyclical and increase with inflation risk, consistent with what consumption-based asset pricing models imply.

Essay 3 examines the predictive power of stock market investor sentiment for Treasury bond returns. Consistent with previous literature, we can identify some in-sample evidence of bond return predictability using sentiment index as predictor. However, this does not generate out-of-sample statistical evidence or economic value for a Bayesian agent. How to translate in-sample evidence of bond return predictability into real-time economic gains remains a challenge.

Keywords: *Bond Return Predictability, Bond Risk Premia, Parameter Uncertainty, Macroeconomic Information, Imperfect Predictors, Predictive Systems, Real Economy, Investor Sentiment, Asset Pricing.*

Contents

List of Tables

List of Figures

Chapter 1

General Introduction

Résumé en Français

Cette thèse de doctorat comprend trois essais en valorisation des actifs financiers, avec une attention particulière sur la prévisibilité du rendement des bons du Trésor Amricain. Ce sujet a reçu beaucoup d'attention dans la littérature récente, mais il reste beaucoup de phénomènes à comprendre.

Dans le premier essai, **Apprentisage en temps réel et la prévisibilité du rendement des bons du Trésor Américain**, co-écrit avec Andras Fulop et Junye LI, nous étudions la preuve statistique et économique de la prévisibilité du rendement des bons en temp réel pour un investisseur bayésien qui se familiarise avec les paramètres, les états et les modèles au fil du temps. Les études précédentes ont decouvert des preuves de cette prévisibilité. Par exemple, Fama et Bliss (1987) et Campbell et Shiller (1991) trouve que linformation dans la courbe de rendement est utile pour prédire les rendements futurs. Cochrane et Piazzesi (2005) trouve que une combination des taux a terme peut expliquer la variation des rendements future avec un R2 de 44%. Toutefois, ces sont des preuve statistiques et ces tests ne sont pas effectué en temps réel. Les investisseurs peuvent etre plus intéressés a savoir sil y a la preuve de la prévisibilité en temps réel. Récemment, Thornton et Valente (2012) trouve que la prévisibilité generé par les taux en terme ne peuvent pas génerer des valeurs économique. Une autre courant de littérature tente didentifier des prédicteur dont la variation se situe en dehors de la courbe de rendement, en particulier certains prédicteurs contenant des informations macroeconomiques. Les études empiriques par Ludvigson et Ng (2009),

Cooper et Priestly (2009), Joslin, Priebsch, et Singleton (2014) et Jiang et Tong (2017), entre autres, montrent que les information macroéconomiques sont utile a prédire rendements futurs. Gargano, Pettenuzzo, et Timmermann (2017) mettent en oeuvre une enquete en utilisant les rendement mensuels et les méthodes MCMC. Ils trouve que les preuves statistiques peuvent se transformer en gains économiques en temps reels. Toutefois, presque toutes ces études utilisent des données macroéconomiques entièrement révisées. Un article récent soutient que la révision des données macroéconomiques peut donner lieu à des preuves fallacieuses de la prévisibilité.

Dans le premier essai, nous revisitons cette question controversée. nous considérons un investisseur bayésien confronté à un problème d'apprentissage. Elle met à jour ses croyances sur les paramètres, les états et les modèles prédictifs au fil du temps. Nous mettons en uvre l'apprentissage bayésien en utilisant l'approche proposée par Fulop et Li (2013). Cet algorithme est générique, efficace et hautement parallèle. Nos données de retour mensuel vont de janvier 1962 à septembre 2017. Nous considérons deux prédicteurs basés sur les taux à terme : le prédicteur de Fama et Bliss (1987) et Cochrane et Piazzesi (2005). Nous construisons cinq prédicteurs basés sur des variables macroéconomiques, dans lesprit de Ludvigson et Ng (2009). nous évaluons la prévisibilité statistique en temps réel en utilisant le R2 par Campbell et Thompson (2008), et évaluons la prévisibilité économique par CERs. Nous avons des résultats intéressants. Nous trouvons des preuves statistiques de la prévisibilité du rendement des obligations en utilisant les informations de la courbe de rendement comme prédicteur. Cependant, ces preuves ne créent pas de valeur économique en temps réel. Lorsque nous utilisons des données macroéconomiques entirement révisées, nous constatons à la fois des preuves statistiques significatives et des gains économiques significatives. Cependant, lorsque nous utilisons des données macroéconomiques en temps réel, ces preuves disparaissent.

Cet article apporte trois contributions à la littérature. Premièrement, nous fournissons un cadre économétrique générique permettant un apprentissage bayésien en temps réel. Deuxièmement, nous attirons l'attention sur l'utilisation d'informations macroéconomiques entièrement révisées par rapport à des informations en temps réel dans les exercices de prévision. Troisièmement, nous constatons que des investissements extrêmes sur le marché obligataire peuvent améliorer les gains économiques à court terme.

Dans le deuxième essai, **Systèmes prédictifs, macro-économie et prévisibilité des rendements obligataires** j'étudie les primes de risque obligataire dans le cadre de systèmes prédictifs. Un grand nombre d'articles précédents sur les primes de risque obligataire se sont concentrés sur la proposition de divers prédicteurs et s'appuyaient fortement sur le modèle de régression linéaire standard. Le modèle est simple et direct, mais il est assez restrictif dans lhypothèse dune relation linéaire parfaite entre les rendements attendus et la valeur actuelle du prédicteur. de plus, en réalit, le prédicteur peut contenir du bruit en raison d'erreurs de mesure potentielles. Lorsque nous utilisons le modle de régression linéaire et un prédicteur potentiellement bruyant, nous pouvons observer des estimations très contre-intuitives des primes de risque obligataire. Par exemple, lorsque nous utilisons le prédicteur de Fama et Bliss (1987), les primes de risque obligataire dans le modèle de régression linéaire montrent 3 pics non raisonnables au milieu des expansions. Une hypothèse est que ces pics inhabituels sont causés par l'utilisation d'un modéle de régression linéaire.

Si tel est le cas, nous devons envisager dutiliser dautres modèles pour produire des primes de risque obligataire plus raisonnables. Dans cet essai, nous utilisons le cadre de système prédictif proposé par Pastor et Stambaugh (2009) pour estimer les primes de risque obligataire. Ce cadre est conu avec une simplicité empirique et un support théorique. D'un côté, ce cadre traite du problème de prédicteur bruyant. En revanche, dans les systèmes prédictifs, la corrélation entre les rendements attendus et inattendus est cruciale pour déterminer les primes de risque obligataire, et nous pouvons imposer des croyances antérieures sur cette corrélation. il s'avère que cela peut affecter de manière substantielle les estimations des primes de risque obligataire.

Dans l'analyse empirique, nous utilisons le gibbs sampling pour estimer les paramètres et les états dans les systèmes prédictifs. Nous rapportons dabord que les systèmes prédictifs fournissent une preuve de prévisibilité plus forte que les modèles de régression linéaire. Pour comprendre les sources économiques de nos résultats, nous étudions comment cette prévisibilité est liée aux variables macroéconomiques. Premièrement, selon le modèle de formation d'habitudes de

Wachter (2006), les primes de risque obligataire devraient augmenter en période de récession, et devrait se déplacer de manière anticyclique. Nous testons donc la corrélation entre plusieurs variables de substitution de la condition macroéconomique et les primes de risque obligataire. Nous trouvons que les estimations des primes de risque obligataire sous des systèmes prédictifs sont anticycliques, alors que certaines estimations issues de modèles de régression linéaire ne le sont pas. deuxièmement, Bansal et Shaliastovich (2013) et Creal et wu (2017) soulignent que le risque d'inflation est un facteur clé des primes de risque obligataire. Nous testons donc la corrélation entre les primes de risque obligataire et proxy du risque d'inflation. Nous montrons que les primes de risque obligataire déduites des systèmes prédictifs augmentent avec le risque dinflation, conformément aux resultats des travaux précédents.

Nous devrions utiliser des systèmes prédictifs pour estimer les primes de risque sur actions ou sur obligations. Les resultats de cet essay et de Pastor et Stambaugh (2009) montrent que les systèmes prédictifs peuvent produire des preuves plus solides de la prévisibilité des rendements des actions ou des obligations.

Dans le troisième essai, **Sentiment des investisseurs et prévisibilité du rendement des obligations**, j'étudie le pouvoir du sentiment des investisseurs boursiers pour prédire les rendements des obligations. Les marchés des actions et des obligations sont deux cibles extrêmement importantes pour les investisseurs, les économistes et les dcideurs. Plusieurs articles ont tenté de comprendre la corrélation entre les rendements des actions et des obligations. Par exemple, Baele, Bekaert, et Inghelbrecht (2010) montrent que la corrélation entre les rendements quotidiens des actions et des obligations pourrait varier dans le temps entre -0,6 et 0,6. Baker et Wurgler (2012) étudient le lien entre les obligations et la section transversale des actions. Leurs résultats suggrent que le sentiment des investisseurs, un prédicteur de la section transversale des rendements boursiers, prédit également les rendements obligataires. Toutefois, cette observation nest pas établie en temps réel et est statistique. Ainsi, dans cet essai, nous testons si l'indice de sentiment des marchès boursiers peut prédire les rendements des obligations en temps rèel. Nous considérons à la fois le modèle de régression linéaire et les systèmes prédictifs.

Nous pouvons identifier certaines preuves dans l'échantillon de la prévisibilité du rendement des obligations en utilisant le sentiment comme prédicteur. Toutefois, lorsque nous utilisons uniquement des informations en temps réel, il nya pratiquement aucune preuve de prévisibilité. Des tests de robustesse supplémentaires donnent des résultats similaires. Nos résultats sont cohérents avec ceux de Thornton et Valente (2012) et Sarno, schneider, et wagner (2016). Comment trouver des preuves en temps réel de la prévisibilité du rendement des obligations reste un défi. Nous nous attendons à voir plus de recherches sur ce sujet.

Introduction

This doctoral thesis investigates several topics in empirical asset pricing, with a focus on Treasury bond return predictability. Treasury bonds not only play an important role in many investors' portfolios, but also attract attention of economists and policymakers. Understanding the risk and return dynamics for this asset has been a long-lasting topic in asset pricing, yet many interesting questions remain to be answered.

In the first essay, **"Real-Time Bayesian Learning and Bond Return Predictability"**, co-authored with Andras Fulop and Junye Li, we study real-time statistical and economic evidence of bond return predictability.

In the literature, a standard way to test evidence of predictability is to run regressions of excess bond returns on some predetermined predictors. Empirical investigations have uncovered some evidence of bond return predictability. For example, 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 a single combination of forward rates can predict excess returns on one- to five-year maturity bonds with R^2 of 0.44 during the period between January 1964 and December 2003. However, a lot of previous 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. More recently, 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 that under affine term structure model framework the evident statistical predictability of bond risk premia hardly turns into investors' economic gains.

Another strand of research try to identify predictors whose variations lie outside the span of yield curve, especially variables containing macroeconomic information. 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 information is included in forecasts of future excess bond returns. 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 (2018) argue that macroeconomic data revision may result in spurious evidence of bond return predictability.

In the first essay, 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 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 using available real-time information 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. In essence, our treatment here is similar to those of Johannes, Korteweg, and Polson (2014), Fulop, Li, and Yu (2015), and Johannes, Lochstoer, and Mou (2016).

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 reply on overlapping excess bond returns at monthly forecasting frequency. Bauer and Hamilton (2017) show that bond returns with overlapping holding-period may induce strong serial correlations in error terms and may raise additional econometric problems when predictors are also persistent. Moreover, Gargano, Pettenuzzo, and Timmermann (2017) point out that some dramatic swings in bond prices can occur over short periods and could be overlooked by using annual overlapping returns. Therefore, similar to Gargano, Pettenuzzo, and Timmermann (2017), we consider one-month holding period and construct monthly non-overlapping excess returns.

We consider two predictors based on forward rates: the forward spreads (FB) of Fama and Bliss (1987) and the forward factor (CP) of Cochrane and Piazzesi (2005), and construct five predictors based on macroeconomic variables: (i) LN, which 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); (2) LNRT1, which is constructed using the historically available real-time macroeconomic vintage data and Bayesian Information Criterion (BIC); (iii) LNRT2, which is simply the first principle component of the historically available real-time macroeconomic vintage data; (iv) LNRT3, which is the first principle component of "first-released" real-time macroeconomic vintage data (Ghysels, Horan, and Moench, 2018); and (v) LNSF, which, similar to Eriksen (2017), is constructed using the forward-looking survey-based macroeconomic data. Given data availability, all empirical tests based on LNSF are performed on quarterly frequency.

We evaluate statistical out-of-sample predictability using the out-of-sample R-squared, 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 Bayesian investor. We obtain 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 translate into investors' economic gains, no matter how strong investors' risk aversion is. 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, LN, we find qualitatively similar results to those of Gargano, Pettenuzzo, and Timmermann (2017). That is, significant statistical evidence of bond return predictability can translate into significant investors' economic gains. However, as discussed by Ghysels, Horan, and Moench (2018), 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 LN is replaced by real-time macroeconomic predictors. We find that whenever the real-time macro factors, LNRT1/LNRT2/LNRT3, 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). Furthermore, when the forward-looking survey-based real-time macro factor, LNSF, is used, we also hardly find any statistical and economic evidence of bond return predictability. This result is different from that found by Eriksen (2017) who uses overlapping returns and ignores the real-time learning. We further show that model combinations do not seem to help uncover significant statistical and economic evidence of bond return predictability whenever the real-time and survey-based macro factors are used, even though some weak evidence can be observed.

Finally, the previous literature has predominantly adopted restrictions on portfolio weights in testing economic evidence. The usual lower and upper weight bounds for risky bonds are -1 and 2, allowing for 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 most of the economic gains quantitatively improve, especially for the short-maturity bonds.

This paper makes three main contributions to the literature. First, we provide a generic econometric framework allowing for 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, in investigating out-of-sample predictability, MCMC algorithms need 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. Therefore, it can be customised to different predictive models easily with comparison to Bayesian MCMC methods that are typically more model-dependent.

Second, we call attention to using fully-revised vs. real-time/survey-based macroeconomic 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. Ghysels, Horan, and Moench (2018) 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.

In the second essay, **"Predictive Systems, Real Economy, and Bond Risk Premia"**, I study bond risk premia in the framework of predictive systems.

A lot of previous papers studying bond risk premia focus on proposing various predictors of bond returns and rely heavily on the standard linear regression model. The model is simple and straightforward, but it is quite restrictive in assuming a perfect linear relationship between the expected return and the current value of the predictor. Moreover, in reality the predetermined predictor can contain noise, due to problems such as measurement error. When we use the traditional linear regression model, combined with a possibly noisy predictor, we may observe very counterintuitive spikes of bond risk premia in expansions. For example, if we use the forward spread from Fama and Bliss (1987) as predictor, the dash-dotted line in Figure C.5 represents the expected return estimates under linear regression model. There are 3 unreasonable spikes occurring in the middle of expansions around 1987, 1992-1995 and 2013-2015 (shaded area represents NBER recessions). One speculation is that these unusual spikes are caused by the use of linear regression model. Therefore we have to consider using other models to produce more economically meaningful bond risk premia.

In this essay, we use the predictive system framework from Pastor and Stambaugh (2009) to estimate bond risk premia. This framework is designed with both empirical simplicity and theoretical support. Pastor and Stambaugh (2009) initially use it to estimate equity risk premia, but it can be easily applied to bond market. On one hand, the framework deals with noisy predictor problem. Under predictive systems, we do not have to follow the common practice in assuming that expected return depends only on the current value of the predictor. Instead, when predictors are noisy or imperfect, expected returns will depend on the history of returns and predictors. On the other hand, within the predictive system framework, the correlation between unexpected returns and expected returns is crucial in determining the risk premium. Per Pastor and Stambaugh (2009), this correlation should be negative for equity returns, because intuitively equity prices tend to fall when discount rates rise. Although stock returns are driven by both cash flow shocks and discount rate news, the latter should have dominant effect. In other words, this negative correlation means that when expected returns experience a positive shock, stock returns will very likely experience a negative shock and prices will decrease. For Treasury bond returns, this correlation could be even more negative, as Treasury bond prices are only subject to discount rate news. We consider three different prior beliefs to incorporate the hypothesis that bond prices tend to fall when discount rates increase: a more informative prior belief, a less informative prior belief, and a noninformative prior belief. Different priors will generate different parameter estimates, and will result in different expected return estimates, as conditional expected return under the predictive system framework is a function of parameters and observation of returns and predictors.

We use the dataset by Gurkaynak, Sack, and Wright (2007) to construct monthly excess bond returns. We use non-overlapping bond returns to avoid the econometric problem pointed out by Bauer and Hamilton (2018). Also, according to Gargano, Pettenuzzo, and Timmermann (2017), using monthly non-overlapping returns, instead of 12-month overlapping returns, can better capture short-term variations in bond risk premia. We consider a wide range of predictors in the literature: the forward spreads (FB) from Fama and Bliss (1987), the combination of forward rates (CP) from Cochrane and Piazzesi (2005), the cycle factor (CF) from decomposition of yield curve in Cieslak and Povala (2015), and the macroeconomic predictor (LN) constructed in the spirit of Ludvigson and Ng (2009).

In the empirical analysis, we use the Gibbs sampling algorithm from Pastor and

Stambaugh (2009) to estimate parameters and states in predictive systems. We first report that predictive systems deliver stronger evidence of predictability than linear regression models. For example, when we use the forward spread as predictor and regress the realized returns on expected returns estimated using predictive systems and linear regression model, the R^2 values from linear regression models are between 1% and 4% whereas the R^2 values from predictive systems are between 3% and 8%. Moreover, if we compare the R^2 values within predictive systems, the specifications imposing negative prior beliefs (negative prior correlations between unexpected and expected returns) generally produce higher R^2 values than the noninformative predictive system.

To understand the economic sources of our findings, we investigate how such predictability is related to macroeconomic variables. First, according to the habit-formation model of Wachter (2006), bond risk premia should increase in recessions because of reduced surplus consumption. So, bond risk premia should move in a countercyclical manner. To test this, we use several proxies of macroeconomic conditions and find that estimates of bond risk premia inferred by predictive systems are countercyclical, whereas some estimates of risk premia from linear regression models do not show such a pattern. Second, the long-run risk models by Bansal and Shaliastovich (2013) and Creal and Wu (2017) point out that inflation risk is a key driver of bond risk premia, and the empirical work of Wright (2011) also suggest inflation risk is an important component of bond risk premia. So we test the correlation between bond risk premia and proxy of inflation risk. Our results show that bond risk premia inferred by predictive systems increase with inflation uncertainty. In sum, predictive systems can generate more economically meaningful dynamics of bond risk premia than standard linear regression models.

We should use predictive systems to estimate equity or bond risk premia. Pastor and Stambaugh (2009) show that predictive systems produce better in-sample fitting for predicting equity returns. Our results show that predictive systems can produce stronger evidence of bond return predictability than simple linear regression models, and the inferred bond risk premia are much more economically reasonable. Whenever we propose a new predictor for equity or bond returns, we should consider using predictive systems to improve the evidence of predictability. Moreover, our results confirm that several macroeconomic variables can be potential economic drivers of time-varying bond risk premia, consistent with the prediction of consumption-based asset pricing models.

In the third essay, **"Investor Sentiment and Bond Return Predictability"**, I study the power of stock market investor sentiment in predicting Treasury bond returns.

Equity and Treasury Bond markets are two extremely important targets for today's investors, economists, and policymakers. The bond market is closely linked with monetary policy, and is expected to interact with the stock market. Several papers have tried to understand the correlation between stock and government bond returns. For example, Baele, Bekaert, and Inghelbrecht (2010) show that correlation between daily equity and bond returns could vary over time between

-0.60 and 0.60. Baker and Wurgler (2012) study the link between government bonds and the cross section of stocks. They observe that stock market sentiment and flights to quality are anecdotally associated over time during special financial market episodes, such as during the recent financial crisis. Their results suggest that sentiment, a predictor of the cross section of stock returns, predicts excess government bond returns. This delivers evidence that the expected returns of stocks and bonds are firmly linked.

However, the above observation of Baker and Wurgler (2012) is drawn on pure in-sample tests. Recently, several studies focus on whether in-sample evidence of bond return predictability could generate economic values for real-time investors. Thornton and Valente (2012) find that using forward spreads as predictor does not create higher utility compared with using expectations hypothesis model, which indicates no predictability. Sarno, Schneider, and Wagner (2016) use affine term structure models and reach similar conclusion. Although Gargano, Pettenuzzo, and Timmermann (2017) find that in-sample evidence is linked with out-of-sample statistical and economic evidence, their tests use fully revised and not real time information. Fulop, Li, and Wan (2018) find that the strong statistical and economic evidence from fully revised macroeconomic data vanishes when real-time or survey-based information is used instead. So, up to now in the literature, there is generally negative evidence of real-time bond return predictability.

In this essay, we test whether the stock market sentiment index can predict government bond returns both in and out of sample. Traditional tests mainly rely on linear regression model, but we also consider the predictive system model proposed by Pastor and Stambaugh (2009). The predictive system was originally used to estimate equity risk premia and was used recently by Wan (2018) to evaluate bond risk premia. In real time analysis, we avoid hindsight problem and use only information available in real time.

Our results show that we can identify some in-sample evidence of bond return predictability. Using stock market sentiment as predictor, we can generate in-sample R-squared of similar magnitude as when we use information from the yield curve. For example, when we use predictive system model with more informative priors, the R^2 values generated using Fama and Bliss (1987) predictor vary from 6.19% to 6.50%, and those generated using sentiment index vary from 6.32% to 6.49%. However, when we switch to use only real-time data, both statistical and economic measures suggest that there is hardly any evidence of predictability. Additional robustness tests deliver similar results. Our conclusion is consistent with what Thornton and Valente (2012) and Sarno, Schneider, and Wagner (2016) have found. How to prove the link between in-sample and out-of-sample predictability still remains a challenge. We expect to see more research on this topic.

Chapter 2

Real-Time Bayesian Learning and Bond Return Predictability*

*This paper is co-authored with Andras Fulop (ESSEC Business School) and Junye Li (ESSEC Business School). We thank seminar participants from Eleventh Annual Risk Management Conference, 2017 Annual Meeting of LACEA-LAMES, 2017 China International Risk Forum (CIRF), 2018 China International Conference in Finance (CICF), City University of Hong Kong, and Workshop on Bayesian Methods in Finance.

2.1 Introduction

The expectations hypothesis (EH) of the term structure of interest rates asserts that the long-term rate is equal to the average of expected future short rates plus a constant risk premium. 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 can not 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 hardly turns into investors' economic gains.

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 contained 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 (2018) 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 that 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. In essence, our treatment here is similar to those of Johannes, Korteweg, and Polson (2014), Fulop, Li, and Yu (2015), and Johannes, Lochstoer, and Mou (2016).

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 overlapping excess bond returns at monthly forecasting frequency. Bauer and Hamilton (2017) show that bond returns with overlapping holding-period may induce strong serial correlations in error terms and may raise additional econometric problems when predictors are also persistent. Moreover, Gargano, Pettenuzzo, and Timmermann (2017) point out that some dramatic swings in bond prices can occur over short periods and could be overlooked by using annual overlapping returns. Therefore, similar to Gargano, Pettenuzzo, and Timmermann (2017), we consider one-month holding period and construct monthly non-overlapping excess returns.

We consider two predictors based on forward rates: the forward spreads (FB) of Fama and Bliss (1987) and the forward factor (CP) of Cochrane and Piazzesi (2005), and construct five predictors based on macroeconomic variables: (i) LN, which 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); (2) LNRT1, which is constructed using the historically available real-time macroeconomic vintage data and Bayesian Information Criterion (BIC); (iii) LNRT2, which is simply the first principle component of the historically available real-time macroeconomic vintage data; (iv) LNRT3, which is the first principle component of "first-released" real-time macroeconomic vintage data (Ghysels, Horan, and Moench, 2018); and (v) LNSF, which, similar to Eriksen (2017), is constructed using the forward-looking survey-based macroeconomic data. Given data availability, all empirical tests based on LNSF are performed on quarterly frequency.

We evaluate statistical out-of-sample predictability using the out-of-sample R-squared, 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 Bayesian investor. We obtain 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 translate into investors' economic gains, no matter how strong investors' risk aversion is. 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, LN, 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 (2018), 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 LN is replaced by real-time macroeconomic predictors. We find that whenever the real-time macro factors, LNRT1/LNRT2/LNRT3, 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). Furthermore, when the forward-looking survey-based real-time macro factor, LNSF, is used, we also hardly find any statistical and economic evidence of bond return predictability. This result is different from that found by Eriksen (2017) who uses overlapping returns and ignores the real-time learning. We further show that model combinations do not seem to help uncover significant statistical and economic evidence of bond return predictability whenever the real-time and survey-based macro factors are used, even though some weak evidence can be observed.

Finally, 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 for 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 most of the economic gains quantitatively improve, especially for the short-maturity bonds.

Our work makes three main contributions to the literature. First, we provide a generic econometric framework allowing for 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, in investigating out-of-sample predictability, MCMC algorithms need 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. Therefore, it can be customised to different predictive models easily with comparison to Bayesian MCMC methods that are typically more model-dependent.

Second, we call attention to using fully-revised vs. real-time/survey-based macroeconomic 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. Ghysels, Horan, and Moench (2018) 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 2.2 presents the predictive models and introduces the Bayesian learning approach. Section 2.3 discusses how we statistically and economically evaluate the predictive performance of each model. Section 2.4 presents the data and summary statistics. Section 2.5 provides the main empirical results and Section 2.6 concludes the paper.

2.2 Bayesian Learning and Bond Return Predictability

2.2.1 Predictive Models

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

$$y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)}, \tag{2.1}$$

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

$$f_t^{(n-m,n)} \equiv p_t^{(n-m)} - p_t^{(n)}, \tag{2.2}$$

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,

$$rx_{t+m}^{(n)} = p_{t+m}^{(n-m)} - p_t^{(n)} - m \cdot y_t^{(m)}, \tag{2.3}$$

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 or one-quarter, 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

$$rx_{t+1}^{(n)} = \alpha + \beta X_t + \epsilon_{t+1}, \tag{2.4}$$

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. Equation (2.3) suggests that $rx_{t+1}^{(n)}$ represents the non-overlapping excess bond return with one-month or one-quarter 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 (2.4), we also introduce the stochastic volatility model, which takes the form of

$$rx_{t+1}^{(n)} = \alpha + \beta X_t + e^{h_{t+1}} \epsilon_{t+1}, \tag{2.5}$$

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

$$h_{t+1} = \mu + \phi h_t + v_{t+1}, \tag{2.6}$$

where h_t is stationary and mean-reverting when $|\phi| < 1$, and $v_t \sim N(0, \sigma_h^2)$. 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, we consider two types of predictors: predictors based on forward rates, i.e., the forward spreads (FB) of Fama and Bliss (1987) and the forward factor (CP) of Cochrane and Piazzesi (2005), and predictors based on different types of macroeconomic variables.

The FB factor is simply defined as:

$$FB_t^{(n,m)} = f_t^{(n-m,n)} - m \cdot y_t^{(m)}. \tag{2.7}$$

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,

$$\overline{rx}_{t+1} = \lambda_0 + \lambda \mathbf{f}_t + u_{t+1}, \tag{2.8}$$

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

$$CP_{t+1} = \hat{\lambda}_0 + \hat{\lambda}\mathbf{f}_{t+1}. \tag{2.9}$$

The macro factors are constructed from a large set of different types of macroeconomic variables using principal component analysis similar to that of Ludvigson and Ng (2009). 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-type factors as follows

$$LN_{t+1} = \hat{\gamma}_0 + \hat{\gamma}\hat{\mathbf{F}}_{t+1}, \tag{2.10}$$

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

$$\overline{rx}_{t+1} = \gamma_0 + \gamma \hat{\mathbf{F}}_t + e_{t+1}. \tag{2.11}$$

More details on construction of different LN-type macro predictors will be discussed in Section 2.4.

We only consider single-predictor predictive models. We christen 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 Equation (2.4), 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 found to be even worse than the standard EH model. Therefore, we report results that take EH-CV as the only benchmark.

2.2.2 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, $y_{1:t} = \{rx_{1:t}^{(n)}, X_{1:t|t}\}$. $X_{1:t|t}$ denotes the time series of the predictor from time 1 to t based on information only available up to time t and suggests that our predictors are updated in real time at each time in the out-of-sample period.

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,

$$p(h_t, \Theta | y_{1:t}, \mathcal{M}_i) = p(h_t | \Theta, y_{1:t}, \mathcal{M}_i) p(\Theta | y_{1:t}, \mathcal{M}_i), \tag{2.12}$$

where $p(h_t | y_{1:t}, \Theta, \mathcal{M}_i)$ solves the state learning (filtering) problem, and $p(\Theta | y_{1:t}, \mathcal{M}_i)$

addresses the parameter learning issue. Bayesian learning or Investor's belief updating therefore corresponds to updating this joint posterior distribution.

For the linear predictive model of Equation (2.4), 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 (2.12) 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, $p(rx_{1:t}^{(n)}|\Theta, \mathcal{M}_i)$. 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 essence, our treatment here is similar to those of Johannes, Korteweg, and Polson (2014), Fulop, Li, and Yu (2015), and Johannes, Lochstoer, and Mou (2016). 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

$$p(rx_{t+1}^{(n)}|y_{1:t},\mathcal{M}_i) = \int p(rx_{t+1}^{(n)}|h_t,\Theta,y_{1:t},\mathcal{M}_i)p(h_t|\Theta,y_{1:t},\mathcal{M}_i)p(\Theta|y_{1:t},\mathcal{M}_i)dh_t d\Theta, \tag{2.13}$$

and an estimate of the marginal likelihood,

$$p(rx_{1:t}^{(n)}|\mathcal{M}_i) = \prod_{s=1}^{t-1} p(rx_{s+1}^{(n)}|y_{1:s},\mathcal{M}_i), \tag{2.14}$$

both of which account for all uncertainties related to parameters and state. Equation (2.14) 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

$$\mathcal{BF}_t \equiv \frac{p(rx_{1:t}^{(n)}|\mathcal{M}_i)}{p(rx_{1:t}^{(n)}|\mathcal{M}_j)} = \frac{p(rx_t^{(n)}|y_{1:t-1},\mathcal{M}_i)}{p(rx_t^{(n)}|y_{1:t-1},\mathcal{M}_j)}\mathcal{BF}_{t-1}, \tag{2.15}$$

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

$$E[\theta_{t+1}|y_{1:t}] = E[E[\theta|y_{1:t+1}]|y_{1:t}] = E[\theta|y_{1:t}] = \theta_t. \tag{2.16}$$

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 and then is moved forward using a Gaussian mixture proposal for the fixed parameters in one block. 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 their 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 B.1 provides two sets of the priors that are used in the paper.

2.2.3 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 (2.14), which can be directly used to combine models.

2.2.3.1 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,

$$p_{SBM}(rx_{t+1}^{(n)}|y_{1:t}) = p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i)\mathbb{I}(\max_{\mathcal{M}} p(rx_{1:t}^{(n)}|\mathcal{M}) = \mathcal{M}_i). \tag{2.17}$$

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

2.2.3.2 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,

$$p_{BMA}(rx_{t+1}^{(n)}|y_{1:t}) = \sum_{i=1}^{N} w_{i,t} \times p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i), \tag{2.18}$$

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

$$p(\mathcal{M}_i | rx^{(n)}_{1:t}) = \frac{p(rx^{(n)}_{1:t} | \mathcal{M}_i) p(\mathcal{M}_i)}{\sum_{j=1}^{N} p(rx^{(n)}_{1:t} | \mathcal{M}_j) p(\mathcal{M}_j)}, \tag{2.19}$$

in which $p(rx^{(n)}_{1:t} | \mathcal{M}_i)$ 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.

2.2.3.3 Equal-weighted Model Average (EMA)

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

$$p_{EMA}(rx^{(n)}_{t+1} | y_{1:t}) = \sum_{i=1}^{N} w_{i,t} \times p(rx^{(n)}_{t+1} | y_{1:t}, \mathcal{M}_i), \tag{2.20}$$

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

2.2.3.4 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,

$$p_{UMA}(rx_{t+1}^{(n)}|y_{1:t}) = \sum_{i=1}^{N} w_{i,t} \times p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_j), \tag{2.21}$$

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

$$p(\mathcal{M}_i|rx_{1:t}^{(n)}) = \frac{CER_{i,t}}{\sum_{j=1}^{N} CER_{j,t}}, \tag{2.22}$$

in which $CER_{i,t}$ is the certainty equivalent return computed using Equation (2.31) for the period from the beginning date of out-of-sample to the current time t.

2.3 Assessing Out-of-Sample Performance

2.3.1 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

$$\widehat{SSE}(t) = \sum_{s=t_0}^{t} (rx_{s+1}^{(n)} - \widehat{rx}_{s+1}^{(n)})^2. \tag{2.23}$$

Furthermore, denote the point forecast from the expectations hypothesis as $\overline{rx}_{t+1}^{(n)}$. Then the SSE for the expectations hypothesis model is given by

$$\overline{SSE}(t) = \sum_{s=t_0}^{t} (rx_{s+1}^{(n)} - \overline{rx}_{s+1}^{(n)})^2. \tag{2.24}$$

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

$$R^2_{OS} = 1 - \frac{\widehat{SSE}(T)}{\overline{SSE}(T)}, \tag{2.25}$$

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 EH forecast.

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 well in terms of power and size properties.

2.3.2 Economic Value and Certainty Equivalence Returns

In evaluating economic predictability, we consider a real-time Bayesian 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},

$$\max_{\omega} \mathbf{E}[U(W_{t+1})|y_{1:t}, \mathcal{M}_i], \tag{2.26}$$

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,

$$U(W_{t+1}) \equiv U(\omega_t^{(n)}, rx_{t+1}^{(n)}) = \frac{W_{t+1}^{1-\gamma}}{1-\gamma}. \tag{2.27}$$

The portfolio value evolves according to

$$W_{t+1} = (1 - \omega_t^{(n)}) \exp(r_t^f) + \omega_t^{(n)} \exp(r_t^f + rx_{t+1}^{(n)}), \tag{2.28}$$

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,

$$\mathbf{E}[U(W_{t+1})|y_{1:t}, \mathcal{M}_i] = \int U(\omega_t^{(n)}, rx_{t+1}^{(n)}) p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i) drx_{t+1}^{(n)}, \tag{2.29}$$

where the predictive distribution of excess bond returns, $p(rx_{t+1}^{(n)}|y_{1:t}, \mathcal{M}_i)$, is given by Equation (2.13).

Our investor is Bayesian. When computing expected utility in Equation (2.29), 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

$$\hat{w}_t^{(n)} = \arg\max \frac{1}{M} \sum_{j=1}^{M} \left\{ \frac{\left[(1 - \omega_t^{(n)}) \exp(r_t^f) + \omega_t^{(n)} \exp(r_t^f + rx_{t+1}^{(n),(j)}) \right]^{1-\gamma}}{1 - \gamma} \right\}. \tag{2.30}$$

The above portfolio weight in Equation (2.30) 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

$$CER = \left(\frac{\sum_{t=1}^{T} \hat{U}_t}{\sum_{t=1}^{T} \bar{U}_t} \right)^{\frac{1}{1-\gamma}} - 1. \tag{2.31}$$

2.4 Data and Summary Statistics

We construct monthly or quarterly frequency 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 overlapping annual excess bond returns in monthly/quarterly forecasting frequency. Bauer and Hamilton (2017) argue that the overlapping bond returns induce strong serial correlations in error terms in predictive regressions, and may raise additional econometric problems when predictors are persistent. Moveover, Gargano, Pettenuzzo, and Timmermann (2017) point out that non-overlapping returns may better reflect short-term swings in bond prices. Therefore, similar to Gargano, Pettenuzzo, and Timmermann (2017), we consider one-month or one-quarter holding period and construct non-overlapping monthly or quarterly excess bond returns. The data span from January 1962 to September 2017.

Ghysels, Horan, and Moench (2018) argue that macroeconomic data revision and publication delay may generate spurious evidence of bond return predictability. Eriksen (2017) find evidence that macroeconomic information extracted from survey of professional forecasts has predictive power for excess bond returns.

Therefore, we employ three types of macroeconomic data: the first is fully-revised, the second is real-time, and the third is forward-looking survey-based, to construct LN-type macro predictors.

We construct our first LN-type macro predictor, LN, using the fully-revised macroeconomic data, downloaded from St. Louis Fed, by relying on the optimal subset of PCs and their powers recommended by Ludvigson and Ng (2009) and Gargano, Pettenuzzo, and Timmermann (2017).[1]

We then construct two LN-type macro predictors, LNRT1 and LNRT2, using the historically available real-time macroeconomic data published by McCracken and Ng (2016).[2] Due to publication delay, 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 LNRT1 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 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 then re-build LNRT1 at each subsequent time. LNRT2 is constructed simply as the

[1]This subset is $[\hat{F}_{1,t}, \hat{F}^3_{1,t}, \hat{F}_{3,t}, \hat{F}_{4,t}, \hat{F}_{8,t}]$. Ludvigson and Ng (2009) choose this subset using BIC and overlapping excess returns. The fully-revised macroeconomic data are downloaded at https://research.stlouisfed.org/econ/mccracken/fred-databases, accessed November 2017.

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

real-time first PC.

The macro predictors, LNRT1 and LNRT2, are extracted from the best macroeconomic information available at each historical point. Similar practice is used by Eriksen (2017) and Giacoletti, Laursen, and Singleton (2018). However, Ghysels, Horan, and Moench (2018) define "real-time" in a different way: real-time value is the "first release" of one macro series for one specific historical month. For instance, the GDP growth of March 1980 released in April 1980 might be different from the GDP growth of March 1980 released in January 2010 due to possible revisions happening between 1980 and 2010 (if we consider a one-month publication delay), and the observation released in April 1980 is called the "first release" for GDP growth of March 1980. We then construct the corresponding first-release-based real-time macro factor, LNRT3, as the first principle component of 63 macroeconomic series, which are the same as those used in Ghysels, Horan, and Moench (2018).[3] Given data availability, the whole sample for LNRT3 is from April 1982 to December 2015.

Finally, we construct a survey-based LN-type macro predictor, LNSF. The survey forecasts data are from Survey of Professional Forecasters. Given data availability, all empirical tests based on survey forecasts data are performed on quarterly frequency from the fourth quarter of 1968 to the fourth quarter of 2014. We collect survey forecasts for six macroeconomic fundamentals that include GDP, the GDP price index, corporate profits after tax, the unemployment rate, indus-

[3]In addition, we build another predictor using Bayesian Information Criterion and the first releases of macro data. However, it performs worse than LNRT3.

trial production, and housing starts. Similar to Eriksen (2017), for each variable we compute one to four quarters ahead expected growth rates relative to forecaster's own nowcast and then aggregate the individual growth forecasts to median forecast. Therefore, there are a total of 24 macroeconomic forecast time series, whose first three PCs are used to build LNSF.

Table B.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 C.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 LN-type macro factors. We find that (1) the 2-year FB is positively skewed, whereas other FBs are negatively skewed, and kurtosis of FBs decrease with respect to maturity. The FB factors are very persistent; (2) the CP factor is positively skewed and leptokurtic, and its persistence is weaker than FBs'; (3) the LN-type macro factors have very different statistical properties.

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

2.5 Empirical Results

2.5.1 Parameter Learning and Sequential Model Comparison

Different from batch estimation methods such as Bayesian MCMC methods, our Bayesian learning approach provides us 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 FB as their only predictor (i.e., FB-SV and FB-CV). Figure C.2 presents the sequential learning of the fixed parameters for FB-SV, and Figure C.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 (5, 95)% 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, suggesting that parameter uncertainties decrease over time.

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 a few 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 (5, 95)% credible interval to sufficiently narrow down.

Third, the last panel of Figure C.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 (5, 95)% 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 C.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 C.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 (2.14). We can then construct the sequential Bayes factors and use them for real-time model analysis and comparison. The last panel of Figure C.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 C.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 outperformance of the stochastic volatility model over the constant volatility model is the strongest on the 5-year excess bond returns before 1980, whereas afterwards, its outperformance becomes the strongest on the 2-year excess bond returns.

2.5.2 Evidence of Bond Return Predictability

Due to availability of real-time macroeconomic vintage data, we set the out-of-sample period from August 1999 to September 2017 in the analysis with FB, CP, LN and LNRT1/LNRT2, and from April 1994 to December 2015 in the analysis with LNRT3; However, when we use LNSF as the predictor, the out-of-sample period is set from the first quarter of 1994 to the fourth quarter of 2014. 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_{OS}^2 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 2.5.2.4. For most part of this section, we use the first set of priors in Table B.1 in our Bayesian learning. We will have a robustness check on sensitivity to priors in the last subsection.

2.5.2.1 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 B.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 to 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 makes no difference 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%, respectively. Third, the results from the models based on CP are different because CP-SV performs better than CP-CV and only those R^2_{OS}'s 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 (2.26). We compute the corresponding certainty equivalent returns (CERs) for each model using formula (2.31), 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 significant. However, we do find that whenever investors become more risk-averse, those CERs for long-maturity (4- and 5-year) 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.

2.5.2.2 Revised, Real-Time, and Forward-Looking Macroeconomic Information

Several studies show that macroeconomic variables contain rich information on future excess bond returns beyond information contained in 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 (2018) argue that macroeconomic data revision and publication delay may introduce spurious bond return predictability. Having taken into account the issue of data revision and publication delay, Eriksen (2017) find evidence that macroeconomic information extracted from survey of professional forecasts has predictive power for future excess bond returns.

To check whether macroeconomic variables contain useful information on future excess bond returns and whether macroeconomic data revision and publication delay make a difference, we construct five LN-type macroeconomic factors as discussed in Section 2.4: LN from fully-revised macroeconomic data; LNRT1 and LNRT2 from historically available real-time macroeconomic data; LNRT3 from the first released macroeconomic data; and LNSF from the survey of professional forecasts.

Panel A of Table B.4 presents R^2_{OS}'s from the models using LN-type macro 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, LN, and stochastic volatility. For example, the R^2_{OS}'s from LN-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. More 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 equal to 5 (Panel C), and for 2-, 3-, and 4-year excess bond return when the coefficient of risk aversion is equal to 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 LN by setting the out-of-sample period ranging from January 1990 to December 2015, which is the same as that used in Gargano, Pettenuzzo, and Timmermamm (2017).

However, we have to take into account the issue of macroeconomic data revision and publication delay in analysis of bond return predictability. For this reason, we implement the same empirical investigation as above using real-time and survey-based LN-type macro factors, LNRT1, LNRT2, LNRT3 and LNSF, instead of LN. We find strikingly different results. No matter which real-time/survey-based macro factors we use, whether we introduce stochastic volatility, or how strong the investor's risk aversion is, there hardly exist either statistical evidence or economic evidence of bond return predictability. It seems that there is evidence that the first principle component extracted from a large panel of real-time macroeconomic data (LNRT2) works better than that constructed using the Bayesian Information Criterion (LNRT1). LNRT3 produces negative statistical R^2_{OS}'s and economic gains are close to zero. In addition, our result based on LNSF is different from that in Eriksen (2017) who finds both statistical and economic evidence of bond return predictability from the survey-based macro factor. The main reason may be because Eriksen (2017) ignores real-time learning and uses overlapping returns.

2.5.2.3 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 four LN-type predictors are constructed completely differently, we implement our model combination schemes for five groups of models, each of which contains only one macro predictor. Panel A of Table B.5 presents the model combination results from using predictors of FB, CP, and LN. We find that the R^2_{OS}'s from BMA, EMA and UMA are in general larger than those in Table B.3 and Table B.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 is equal to 5, and the 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 is equal to 10.

Panels B and C presents the model combination results from using predictors of FB and CP, together with LNRT1 and LNRT2, 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 some statistical evidence in Panel C from SBM and UMA for 2-year bond returns when LNRT2 is used. Whenever we combine models based on FB, CP and LNRT3 in Panel D, we find some significant statistical evidence from SBM and EMA for all-maturity bond returns. However, none of these model combination schemes produces any significant economic gains, no matter how strong the investor's risk-aversion is.

Panel E presents model combination results from using FB, CP and LNSF based on quarterly data. Similarly, we can not see any obvious statistical and economic evidence of bond return predictability. However, it seems that for 4- and 5-year bond returns, EMA produces some (weak) statistical evidence and economic evidence when investor becomes more risk-averse.

2.5.2.4 Extreme Investments and Predictive Performance

Up to now, we have restricted portfolio weight in between -1 and 2 when we implement asset allocation. However, different from the equity market, it may be feasible for investors to take extreme positions in the bond market, facilitated by Repo agreements for instance. Therefore, in this part, we allow investors to take their investment decision without any restrictions on portfolio weights.

Table B.6 presents CERs from all individual models with the coefficient of risk aversion equal 5. We find that with comparison to those in Table B.3 and Table B.4, nearly all CERs greatly improve, especially those for short-maturity bonds (2-year and 3-year). This suggest that without imposing any restrictions on portfolio weights, 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.

However, we still find that CERs from FB- and CP-based models are hardly statistically significant. The CERs from LN-based models (LN-CV and LN-SV) for 2-year bond returns become much higher and highly statistically significant. Among all models based on real-time/survey-based macro factors, CERs from LNRT2-CV is 1.51% and statistically significant for 2-year bond returns and both LNSF-CV and LNSF-SV produce positive CERs (1.03% and 2.31%), which are marginally statistically significant. Similar results are also observed when we set the coefficient of risk aversion equal to 3 or 10.

It seems a new result in literature on bond return predictability and contrast with the general finding on equities where constraints on portfolio weights tend to improve predictive performance in 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.

Table B.7 presents CERs from model combinations with the coefficient of risk aversion equal to 5. As before, we consider the same five groups of models. With comparison to Table B.5 and Table B.6, most of CERs for each maturity improve greatly, especially for 2-year bond. When models based on LN 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 B.5, such significance vanishes when the real-time macro factor, LNRT1/LNRT2/LNRT3, is grouped together with FB and CP (Panel B, C, and D). In panel E, when combining models based on FB, CP, and LNSF, we find that the CERs from EMA are marginally significant for all-maturity bonds, whereas CERs from the other three combination schemes are not statistically significant. We find qualitatively similar results when the coefficient of risk aversion is equal to 3 or 10.

2.5.2.5 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 B.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.

2.6 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 five 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), three are constructed using the real-time macroeconomic panel data, and one is constructed from forward-looking survey forecasts on macroeconomic variables. We compare the predictive performance of expectations hypothesis and predictive models. Statistical out-of-sample predictability is evaluated using the out-of-sample R-squared, R^2_{OS}, of Campbell and Thompson (2008), whereas economic out-of-sample predictability is evaluated 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). Also, Gargano, Pettenuzzo, and Timmermann (2017) point out that some dramatic swings in bond prices can occur over short periods lasting less than a year, and could be overlooked by using annual overlapping returns. Therefore, similar to Gargano, Pettenuzzo, and Timmermann (2017), we mainly consider one-month and one-quarter holding period and construct non-overlapping excess bond returns. We find some statistical evidence using information contained in forward rates. However, such statistical predictability can not generate any economic value for investors. Furthermore, strong statistical and economic evidence from fully revised macroeconomic data vanishes when real-time and survey-based macroeconomic information is used. We also show that extreme investments in bonds could improve short-run bond return predictability.

Chapter 3

Predictive Systems, Real Economy, and Bond Risk Premia*

*I am grateful to my supervisors, Andras Fulop and Junye Li. I also thank Roméo Tédongap and Elise Gourier for helpful comments.

3.1 Introduction

Understanding time-variation in U.S. Treasury bond risk premia is one of the most discussed topics in asset pricing. A lot of previous papers focus on proposing various predictors of bond returns and rely heavily on the standard linear regression model. The model is simple and straightforward, however it is restrictive in assuming a perfect linear relationship between the expected return and the current value of the predictor. Moreover, in reality the predetermined predictor can contain noise, due to problems such as measurement error. Sometimes we observe very counterintuitive variation of bond risk premia if we use the traditional linear regression model. For example, if we use the forward spread from Fama and Bliss (1987) as predictor, the dash-dotted line in Figure C.5 represents the expected return estimates under linear regression model. There are 3 unreasonable spikes occurring in the middle of expansions around 1987, 1992-1995 and 2013-2015 (shaded area represents NBER recessions). One speculation is that these unusual spikes are caused by the use of linear regression model. Therefore we have to ask: can we use other models to produce more economically reasonable bond risk premia?

In this paper, we use the predictive system framework from Pastor and Stambaugh (2009) to estimate bond risk premia. This framework is designed with both empirical simplicity and theoretical support. Pastor and Stambaugh (2009) initially use it to estimate equity risk premia, but it can be applied to bond market. On one hand, the framework deals with noisy predictor problem. Under predictive systems, we do not have to follow the common practice in assuming that expected return depends only on the current value of the predictor. Instead, when predictors are noisy or imperfect, expected returns will depend on the history of returns and predictors. On the other hand, within the predictive system framework, the correlation between unexpected returns and expected returns is crucial in determining the risk premium. Per Pastor and Stambaugh (2009), this correlation should be negative for equity returns, because intuitively equity prices tend to fall when discount rates rise. Although stock returns are driven by both cash flow shocks and discount rate news, the latter should have dominant effect. In other words, this negative correlation means that when expected returns experience a positive shock, stock returns will very likely experience a negative shock and prices will decrease. For Treasury bond returns, this correlation could be even more negative, as Treasury bond prices are only subject to discount rate news. We consider three different prior beliefs to incorporate the hypothesis that bond prices tend to fall when discount rates rise: a more informative prior belief, a less informative prior belief, and a noninformative prior belief. Different priors will generate different parameter estimates, and will result in different expected return estimates, as conditional expected return under the predictive system framework is a function of parameters and observation of returns and predictors.

We use the dataset by Gurkaynak, Sack, and Wright (2007) to construct monthly excess bond returns. We use non-overlapping bond returns to avoid the econometric problem pointed out by Bauer and Hamilton (2018). Also, using monthly non-overlapping returns, instead of 12-month overlapping returns, can better capture short-term variations in bond risk premia. We consider a wide range of predictors in the literature: the forward spreads (FB) from Fama and Bliss (1987), the combination of forward rates (CP) from Cochrane and Piazzesi (2005), the cycle factor (CF) from decomposition of yield curve in Cieslak and Povala (2015), and the macroeconomic predictor (LN) constructed in the spirit of Ludvigson and Ng (2009).

In the empirical tests, we use the Gibbs sampling algorithm from Pastor and Stambaugh (2009) to estimate parameters and states in predictive systems. We first report that predictive systems generate stronger evidence of predictability than linear regression models. For example, when we use the forward spread as predictor and regress the realized returns on expected returns estimated using predictive systems and linear regression model, the R^2 values from linear regression models are between 1% and 4% whereas the R^2 values from predictive systems are between 3% and 8%. Moreover, if we compare the R^2 values within predictive systems, the specifications imposing negative prior beliefs (negative prior correlations between unexpected and expected returns) generally produce higher R^2 values than the noninformative predictive system.

To understand the economic sources of our findings, we investigate the extent to which such predictability is related to macroeconomic variables. First, according to the habit-formation model of Wachter (2006), bond risk premia should increase in recessions because of reduced surplus consumption. So, bond risk premia should move in a countercyclical manner. To test this, we use several proxies of macroeconomic conditions and find that estimates of bond risk premia inferred by predictive systems are countercyclical, whereas some estimates of risk premia from linear regression models do not show such a pattern. Second, the long-run risk models by Bansal and Shaliastovich (2013) and Creal and Wu (2017) point out that inflation risk is a key driver of bond risk premia, and the empirical work of Wright (2011) also suggest inflation risk is an important component of bond risk premia. So we test the correlation between bond risk premia and proxy of inflation risk. Our results show that bond risk premia inferred by predictive systems rise with inflation uncertainty. In sum, predictive systems can generate more economically meaningful dynamics of bond risk premia than standard linear regression models.

In addition, we also perform out-of-sample tests and report the results in Appendix. We assume a real-time Bayesian investor who takes parameters, latent states, and predictive models as unknown and updates her beliefs using Bayes' rule with respect to information accumulation. She computes the predictive return distribution at each time and maximizes her expected utility by taking into account all relevant uncertainties. Real-time results show that predictive systems generally produce stronger statistical evidence than linear regression models, but such evidence is weaker than in full-sample analysis. This is consistent with the findings of Cieslak and Povala (2015), Haddad and Sraer (2018), and Farmer, Schmidt, and Timmermann (2018) that predictability evidence may be weakened in real time. Moreover, such evidence does not translate into economic gains for the Bayesian investor. Previous studies (Goyal and Welch (2008), Thornton and Valente (2012), Sarno, Schneider, and Wagner (2016), etc) also identify the contrast between statistical and economic metrics in evaluating model performance. Furthermore, we find many more procyclical risk premia when using real-time forecasts of expected returns. This indicates a difference between full-sample and out-of-sample results and imply that full-sample estimates of bond risk premia are more accurate.

We should use predictive systems to estimate equity or bond risk premia. Pastor and Stambaugh (2009) show that predictive systems produce better in-sample fitting when predicting equity returns. Our results show that predictive systems can produce stronger statistical evidence of bond return predictability than simple linear regression models, and the inferred bond risk premia are much more economically reasonable. Whenever we propose a new predictor for equity or bond returns, we should think of using predictive systems to improve the predictability evidence. Moreover, our results confirm that several macroeconomic variables can be potential economic drivers of time-varying bond risk premia, consistent with the prediction of consumption-based asset pricing models.

Literature Review. Our paper relates to the long-lasting literature on time-varying bond risk premia. A lot of papers rely on standard linear regression models and focus on proposing different predictors: Starting from Fama and Bliss (1987), some papers use information within the term structure of interest rates to capture bond return variation (Campbell and Shiller (1991), Cochrane and Piazzesi (2005), etc). Another series of papers use macroeconomic information to explore predictability evidence as Campbell and Cochrane (1999) theorize that bond investors must be compensated for risks associated with macroeconomic activity. Ludvigson and Ng (2009) employ methods of dynamic factor analysis to extract information from a large panel of macroeconomic fundamentals and find that using some principal components as predictor can generate marked countercyclical risk premia. Wright (2011) use international evidence to show that inflation risk is an important component of bond risk premia as the largest risk premia declines occurred in countries that made radical changes in monetary policy, such as introducing inflation targeting. Eriksen (2017) constructs a proxy of expected business conditions using forward-looking survey forecasts and shows that the inclusion of this proxy in standard predictive regressions improves forecast performance both in and out of sample. Cieslak and Povala (2015) decompose yield curve into long-run inflation trend, short-term monetary policy expectations, and a cycle/risk premium factor, and the last significantly captures the movement of excess bond returns. Besides, some recent papers propose other types of variables with different theoretical support. Greenwood and Voyanos (2014) show empirically that the supply and maturity structure of government debt alter the price of duration risk and that the maturity-weighted-debt-to-GDP ratio is positively related to future bond returns. Hanson (2014) argue that when the risk tolerance of bond investors is limited in the short run, fluctuations in mortgage-backed-securities duration can generate significant variation in bond risk premia. Haddad and Sraer (2018) proposes a measure of banks' net exposure to interest rate risk and find that the intermediary sector is compensated for bearing such risk.

Another stream of literature use Affine Term Structure Models (ATSM) to model Treasury bond yields and expected returns. Starting from Duffee (2002) and Dai and Singleton (2002), numerous papers have tried to study bond risk premia under ATSM framework. In particular, Joslin, Priebsch, and Singleton (2014) develop dynamic term structure models with unspanned macro risk and the output and inflation risks accounted for a large portion of the variation in risk premia. Besides, some studies try to use consumption-based asset pricing models to estimate bond risk premia. Among these, Wachter (2006) proposes a habit-formation model which accounts for many features of the nominal term structure of interest rates. Bansal and Shaliastovich (2013) show that bond risk premia rise with inflation risk and fall with real growth risk, then they use a long-run risk model to account for bond return predictability and violation of uncovered interest rate parity in currency market.

Our work also relates to the literature on how prior beliefs or certain economic constraints can generate different asset pricing implications. Some previous studies on return predictability allow for parameter uncertainty, but assume non-informative priors (Stambaugh (1999), Barberis (2000), etc). Recently, Shanken and Tamayo (2012) considers a range of prior beliefs about the risk-return trade-off and the degree of predictability. They find that these beliefs can significantly impact the Bayesian investor's economic gains. Pettenuzzo, Timmermann, and Valkanov (2014) impose positive sign on equity risk premia and certain bounds on conditional Sharpe ratio, and find that such constraints improve both statistical and economic forecast performance. Because the Bayesian methodology can easily incorporate prior beliefs on certain important parameters, it is widely used in the above literature and in estimation of state-space models.

The rest of the paper is organized as follows. Section 3.2 and 3.3 describe variables, models, and data we use. Section 3.4 presents full-sample evidence of predictability. Section 3.5 discusses the link between our expected return estimates and real economy. Section 3.6 concludes.

3.2 Models and Predictors

3.2.1 Predictive Regressions vs. Predictive Systems

We define the log-yield of an n-year bond as $y_t^{(n)} \equiv -\frac{1}{n}p_t^{(n)}$, where $p_t^{(n)} \equiv \ln P_t^{(n)}$, and $P_t^{(n)}$ is the nominal price of an n-year zero-coupon bond at time t. The forward rate is $f_t^{(n-m,n)} \equiv p_t^{(n-m)} - p_t^{(n)}$, and the excess return of an n-year bond is the return from buying an n-year bond at time t and selling it m-period later in excess of the yield on a m-period T-bill rate at time t,

$$rx_{t+m}^{(n)} = p_{t+m}^{(n-m)} - p_t^{(n)} - m \cdot y_t^{(m)}, \tag{3.1}$$

where m is the holding period measured in years and $y_t^{(m)}$ is the annualized T-bill rate. Throughout this paper, we assume n can be 2, 3, 4, or 5 years, and m is one-month, so in the following part we replace $rx_{t+m}^{(n)}$ with $rx_{t+1}^{(n)}$.

The standard way of testing bond return predictability uses the linear regression model:

$$rx_{t+1}^{(n)} = \alpha + \beta X_t + \epsilon_{t+1}, \tag{3.2}$$

where X_t is the predetermined predictor, $\epsilon_t \sim N(0, \sigma_{rx}^2)$ is the error term with a zero-mean normal distribution, and the coefficients α, β, and σ_{rx} are unknown parameters. $rx_{t+1}^{(n)}$ represents the non-overlapping excess bond return with one-month holding period.

Equation (3.2) seems too restrictive to assume a perfect linear relationship between the conditional expected return and the predictor, as in reality predictors are very likely to contain noise and therefore can not deliver the expected return perfectly. Therefore Pastor and Stambaugh (2009) propose the following predictive system to describe the relationship between the predictor and the expected return:

$$rx_{t+1}^{(n)} = \mu_t + u_{t+1} \tag{3.3}$$

$$X_{t+1} = (1 - A)E_X + AX_t + v_{t+1} \tag{3.4}$$

$$\mu_{t+1} = (1 - B)E_{rx} + B\mu_t + w_{t+1} \tag{3.5}$$

where μ_t denotes the latent data generating process of conditional expected returns. Both μ_t and the predictor X_t follow an AR(1) process. The noise terms

$\{u_t, v_t, w_t\}$ are assumed to be distributed identically and independently across t with a covariance Σ,

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma \equiv \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \sigma_v^2 & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix}\right) \tag{3.6}$$

Under the predictive system framework, the predictor X_t can directly impact the expected return μ_t through the correlation ρ_{vw}. Assuming we use one predictor, the predictor is "perfect" (i.e., the conditional expected return is a linear function of this predictor) only when $A = B$ and $\rho_{vw} = \pm 1$. The other covariance parameter ρ_{uw} is of great importance for the whole system as we could incorporate some economically motivated beliefs through this parameter. Henceforth, we follow Pastor and Stambaugh (2009) and refer to ρ_{uw} as the "correlation between expected and unexpected returns", a slightly inaccurate but much simpler description. A negative ρ_{uw} means that unanticipated positive shocks in expected returns tend to be accompanied by unexpected negative returns, therefore asset prices will fall when discount rates increase, which is exactly in line with real-life observation. Pastor and Stambaugh argue that although stock returns are subject to both cash flow shocks and discount rate shocks, the latter will have the dominant effect, therefore ρ_{uw} should be negative. But if we apply similar theory to bond returns, ρ_{uw} is very likely to be close to -1, since the nominal cash flows of Treasury bonds are fixed and the price variation is driven only by discount rate shocks. In order to show how this intuition will affect our results, we follow Pastor and Stambaugh and use three different priors for ρ_{uw}: a more informative prior which restricts ρ_{uw} to lie largely below -0.5, a less informative prior which restricts ρ_{uw} to lie largely below 0, and a noninformative prior which does not restrict the distribution of ρ_{uw} at all. We can interpret these priors as aggressive, less aggressive, and noninformative. More details on the priors and estimation of the predictive system will be shown in Section 3.4.1 and A.4.2.

Pastor and Stambaugh (2009) show that conditional expected returns from the predictive system will be a weighted average of past returns and predictors,

$$E(rx_{t+1}|I_t) = E_{rx} + \sum_{s=0}^{\infty}(\lambda_s \varepsilon_{t-s} + \varphi_s' v_{t-s}), \tag{3.7}$$

where $\varepsilon_t = r_t - E(r_t|I_{t-1})$, v_t is the noise term in Equation (3.4), $I_t = \{rx_{1:t}^{(n)}, X_{1:t}\}$ ($rx_{1:t}^{(n)}$ and $X_{1:t}$ denote the time series of returns and predictors from time 1 to t), and λ and φ are functions of time and parameters in equations (3.3), (3.4), and (3.5) (more details on these equations are in the Appendix). Intuitively, different priors on ρ_{uw} will result in different conditional expected returns, and will further influence evidence of predictability under different specifications.

3.2.2 Predictors

Empirical studies have proposed numerous predictors and we choose the most representative ones with monthly data frequency. Fama and Bliss (1987) find that the forward spread has predictive power for excess bond returns and that its power increases with the forecasting horizon. Cochrane and Piazzesi (2005) show that a combination of forward rates can capture more than 30% of the variation of excess bond returns over the period from January 1964 to December 2003. Cieslak and Povala (2015) find that a risk-premium factor constructed from yield curve decomposition can forecast bond returns across different maturities. Furthermore, Ludvigson and Ng (2009) uncover principal components from a large set of fully revised macroeconomic variables and show that these factors have predictive power for future excess bond returns.

Therefore, we consider the following predictors: the forward spreads (FB) from Fama and Bliss (1987), the forward rates factor (CP) from Cochrane and Piazzesi (2005), the risk-premium/cycle factor (CF) from Cieslak and Povala (2015), and Ludvigson and Ng (2009) predictor is constructed using a panel of macroeconomic variables as Ghysels, Horan, and Moench (2018). In our full-sample analysis, we use fully revised macroeconomic information to construct CF and LN predictor, whereas in our out-of-sample tests we use strictly real-time information to avoid hindsight problem and reconstruct CP, CF, and LN at each time. More details on predictors will be discussed in Section 3.3.

The FB factor is the forward spread:

$$FB_t^{(n,m)} = f_t^{(n-m,n)} - m \cdot y_t^{(m)}. \tag{3.8}$$

As our forecasting frequency is monthly, we have $FB^{(2,1/12)}, FB^{(3,1/12)}, FB^{(4,1/12)}$, and $FB^{(5,1/12)}$ to forecast $rx^{(2)}, rx^{(3)}, rx^{(4)}$, and $rx^{(5)}$, respectively.

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

$$\overline{rx}_{t+1} = \iota_0 + \iota \mathbf{f}_t + u_{t+1}, \tag{3.9}$$

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 constructed as

$$CP_{t+1} = \hat{\iota}_0 + \hat{\iota}\mathbf{f}_{t+1}. \tag{3.10}$$

We construct the cycle factor (CF) following Cieslak and Povala (2015). First we compute trend inflation from: $\tau_t^{CPI} = (1-\upsilon)\sum_{i=0}^{t-1} \upsilon^i \pi_{t-i}$, where $\upsilon = 0.987$, $\pi_t = ln(CPI_t) - ln(CPI_{t-1})$, and we truncate the sum at $N = 120$.[1] Then we regress yields from 1-year to 20-year maturity on trend inflation and label each residual series as maturity-specific cycle $c_t^{(n)}$. We define $\bar{c}_t \equiv (1/19)\sum_{i=2}^{20} c_t^{(i)}$ and regress average excess bond return on $c^{(1)}$ and $\bar{c}$,

$$\overline{rx}_{t+1} = l_0 + l_1 c_t^{(1)} + l_2 \bar{c}_t + u_{t+1}, \tag{3.11}$$

and the CF factor is computed as,

$$CF_{t+1} = \hat{l}_0 + \hat{l}_1 c_t^{(1)} + \hat{l}_2 \bar{c}_t. \tag{3.12}$$

The LN factor is constructed from a large set of macroeconomic variables using principal component analysis similar to that of Ludvigson and Ng (2009). We use only the first principal component (PC) extracted from the macroeconomic panel,

[1]We choose values for the hyper-parameters N and υ, and use truncation following Cieslak and Povala (2015). They show that the results are insensitive to the choice of these numbers. Also, we follow them and use the full term structure in creating $\bar{c}$.

as Ghysels, Horan, and Moench (2018) and Huang, Jiang, and Tong (2018) point out that the first PC has the strongest predictive power.[2] We regress average excess return on the first PC,

$$\overline{rx}_{t+1} = \gamma_0 + \gamma \widehat{PC1}_t + u_{t+1}, \tag{3.13}$$

and the LN predictor is computed as,

$$LN_{t+1} = \hat{\gamma_0} + \hat{\gamma} \widehat{PC1}_{t+1}, \tag{3.14}$$

In this paper we only consider single-predictor specifications. The benchmark model for our tests is the expectations hypothesis (EH), where no predictor is used, or $\beta = 0$ in Equation (3.2).

3.3 Data and Summary Statistics

Some prior studies (Cochrane and Piazzesi (2005), Ludvigson and Ng (2009), etc) use overlapping 12-month excess bond returns in monthly/quarterly forecasting frequency. Recently, Bauer and Hamilton (2018) show that overlapping bond returns will induce strong serial correlations in regression residuals, and may raise additional econometric problems when predictors are persistent. Also, Gargano, Pettenuzzo, and Timmermann (2017) argue that some dramatic swings in bond returns may happen over short periods of time and non-overlapping returns can better capture such short-term variations. Therefore, we construct monthly yields

[2] We also use Bayesian Information Criterion (BIC) to choose the optimal PC structure from each macro data (Bai and Ng (2002), McCracken and Ng (2016)), but the produced predictors generally do not deliver stronger evidence of predictability than the ones using only the first PC. This subject may also raise future research questions on how to choose the optimal PC structure.

on US zero-coupon bonds with maturity 2-, 3-, 4-, and 5-year using the updated dataset of Gurkaynak, Sack, and Wright (2007, GSW henceforth).[3] We consider monthly non-overlapping excess returns, because Bauer and Hamilton (2018) show that using 12-month overlapping returns can produce strong serial correlations in error terms and may cause spurious regression results when predictors are also persistent. Gargano, Pettenuzzo, and Timmermann (2017) argue that the inference problems pointed out by Bauer and Hamilton (2017) largely disappear when using one-month nonoverlapping returns. Moreover, monthly non-overlapping returns can better reflect short-term variations in bond risk premia, as some dramatic swings in bond prices occur over short periods of time lasting less than a year (Gargano, Pettenuzzo, and Timmermann (2017)). Our full sample is from April 1982 to December 2015. Numerous papers (Taylor (1999), Clarida, Gali, and Gertler (2000), etc) point out that the Federal Reserve changed its monetary policy during the early 1980s. Joslin, Priebsch, and Singleton (2014) choose the starting date to be 1985 to make sure this is well after the implementation of new operating procedures.

We use GSW dataset to construct FB and CP predictors. For CF predictor, we first construct trend inflation using Consumer Price Index from FRED database of St. Louis Fed, and we regress yields (also from GSW data) across the term structure on trend inflation to get maturity specific cycle factor. Then we regress

[3]Previous studies such as Cieslak and Povala (2015) and Gargano, Pettenuzzo, and Timmermann (2017) have discussed that different datasets can generate almost identical yields. We refer more detailed discussions to those papers.

average excess return on one-year cycle and mean cycle, and the fitted value is the CF predictor. For LN, we select the most representative macroeconomic variables as Ghysels, Horan, and Moench (2018) and extract the first principal component[4].

Table B.8 presents summary statistics for full-sample excess bond returns and predictors. Panel A shows that both mean and standard deviation of the monthly excess returns increase with respect to maturity, whereas skewness, kurtosis, and first-order autocorrelations decrease with maturity. The mean and standard deviation for the 2-year return are about 0.19% and 0.63%, whereas they increase to 0.36% and 1.61% for the 5-year bond. The skewness and kurtosis for the 2-year excess bond returns are 0.36 and 4.61, respectively, whereas they decrease to 0.03 and 3.52 for the 5-year returns. Gargano, Pettenuzzo, and Timmermann (2017) show that the autocorrelations for 12-month overlapping returns can be over 0.93, but the first-order autocorrelations of 1-month non-overlapping returns in the last row of Panel A are much lower and range from 0.12 (5-year) to 0.22 (2-year).

Panel B presents full-sample summary statistics for the predictors: FB, CP, CF, and LN predictors. We can see that the FB, CP, and CF factors are highly persistent with first-order correlations over 0.9, whereas AC(1) for full-sample LN predictor is 0.81. Panel C reports the correlation matrix of predictors. FB, CP, and CF predictors are all positively correlated with each other and the coefficients are above 0.37. These three predictors represent information from the yield curve. LN is only slightly correlated with other predictors because this macroeconomic

[4]Ghysels, Horan, and Moench (2018) point out that these variables cover largely the same economic categories as the original LN data.

predictor contains information from outside of the yield curve.

3.4 Evidence of Predictability

3.4.1 Bayesian Estimation and Prior Beliefs

In our analysis, we employ the Gibbs sampling by Pastor and Stambaugh (2009) to estimate the parameters and state in the predictive systems. We impose economically motivated prior beliefs on the key parameter of interest, ρ_{uw}: we plot the three priors in Figure C.4. The "more informative" prior is specified such that the prior distribution of ρ_{uw} has nearly 99.5% of its mass below -0.73, with a mean of -0.90. The "less informative" prior is specified such that the prior has nearly 99.5% of its mass below -0.043, with a mean of -0.65. The "noninformative" prior is specified such that the prior is flat on most of the (-1, 1) range, with a mean of 0. Note that in this paper whenever we say "more informative", "less informative", or "noninformative", it is meant for the key parameter ρ_{uw}. Intuitively, putting different priors on ρ_{uw} while keeping unrestricted priors for other parameters in Σ requires a hyperparameter approach. Pastor and Stambaugh choose the prior beliefs on ρ_{uw} according to findings from previous papers on equity research, such as Campbell (1991). But intuitively the negative correlation should be more evident for bond returns because bond prices are only subject to discount rate shocks. Therefore, our choice of the more and less informative priors are slightly more aggressive than the ones by Pastor and Stambaugh, just to account for the possibility of a more negative ρ_{uw}.[5] We can see that the prior distributions for ρ_{uv} and ρ_{vw} in Figure C.4 are largely noninformative between -1 and 1. In addition, Pastor and Stambaugh point out that the expected return process is likely to be persistent, we follow their practice and assume a slightly restrictive prior on B, which is $B \sim N(0.99, 0.15^2)$. It would be interesting to see how information from the data interacts with these prior distributions as the Bayesian agent will update beliefs with new information. The prior distributions on all other parameters are noninformative. More details are provided in the Appendix.

3.4.2 Empirical Results

We assess the degree of predictability by adjusted R-squared (R^2). For the linear regression model we regress the excess returns on the predictor; for the predictive system, we regress excess returns on the expected return μ_t which is filtered by Gibbs sampling. A positive R^2 value suggests evidence of predictability and a higher R^2 indicates that more variance of returns is explained by the independent variable. Pastor and Stambaugh (2009) compute the ratio of R^2 from the system to the R^2 from the regression as $Var[E(\mu_t|I_t)]/Var[E(\mu_t|X_{1:t})]$. This ratio is always above one because $X_{1:t} \subset I_t \equiv \{rx^{(n)}_{1:t}, X_{1:t}\}$. In other words, because the system uses more information, the estimates of μ_t from the predictive system should be at least as precise as the estimates from the linear regression model.

We report results in Panel A of Table B.9. In each block, we compare the R^2

[5] We also perform tests using exactly the same prior distributions as in Pastor and Stambaugh (2009). Our main results remain the same.

values from predictive regressions and predictive systems. The predictive systems are labelled by their prior beliefs about ρ_{uw}: whenever we say "More Informative", "Less Informative", or "Noninformative", it is meant for the key parameter ρ_{uw}.

We have some interesting findings. First, using linear regression models, FB and CP produce the highest R^2 values of all predictors, ranging from 1.93% for the 5-year return to 4.34% for the 2-year return. This highlights the importance of information from the term structure in explaining bond return variations. R^2 values from LN are between 1.66% and 3.72%, confirming that bond returns are linked to macroeconomic information. Also, R^2 values from linear regression models decrease with return maturity, except for CF predictor.

Second, using the same predictor, predictive systems produce much higher R^2's than the corresponding predictive regression. This is not surprising as the estimates of expected returns from the predictive system should be at least as precise as the OLS estimates. Comparing among predictors, CF generates the highest R^2 values using predictive systems, ranging from 8.99% for the 2-year return to 5.89% for the 5-year return. Also, in general the more informative and less informative priors can generate higher R^2's than the noninformative one. This indicates that it is useful to incorporate prior beliefs about the negative correlation between expected and unexpected returns.

In Panel B, we plot the correlation matrix of expected returns for the same predictor under different predictive models. To save space, we report results for 3-year excess returns. Across different predictors, the highest correlation (ranging from 0.83 to 0.96) comes from estimates for more informative and less informative systems, i.e., when the negative priors are effective. Other correlations are much lower, and the correlation between expected return forecasts from linear regression model and more informative predictive system is only between 0.07 and 0.48.

We plot the full-sample posterior distributions for ρ_{uw}, ρ_{uv}, and ρ_{vw} in Figure C.4. In this example, FB is used to predict 3-Year excess returns. First, posterior distributions of ρ_{uw} from the two informative priors are still largely negative. The posterior distribution of ρ_{uw} from the noninformative specification is extremely wide, and its majority lies between 0 and 1, showing a high degree of uncertainty regarding this correlation. Second, consistent with Pastor and Stambaugh (2009), the data should be quite informative about ρ_{uv} because both return and predictor are observable. The three posterior distributions of ρ_{uv} are almost identical, lying between -0.5 and -0.2. Third, we find very different estimates of ρ_{vw}. This parameter describes how much the predictor can directly impact the expected return estimates: the more informative prior produces highest ρ_{vw}, indicating more information in return process is used when we estimate expected returns. As we use less aggressive priors, the center of the distribution moves towards zero. These correlations are well below 1. This confirms that the predictor is noisy and imperfectly correlated with μ_t.

3.5 Link to Real Economy

In Figure C.5, we plot estimates of expected returns from different specifications. We use the FB factor to predict 3-year returns. The shaded area represents NBER recessions. The expected return estimates from the linear regression model seem too volatile to be plausible and there are 3 counterintuitive spikes occurring in expansions around 1987, 1992-1995 and 2013-2015. The bond risk premia from predictive systems seem more reasonable as they increase in recessions. Some previous theoretical work also point out that bond risk premia should move in a countercyclical manner. The habit-formation model by Wachter (2006) show that reduced surplus consumption in bad times will result in higher bond risk premia. To evaluate the cyclical pattern of bond risk premia, we test its correlation with several proxies of macroeconomic conditions. The proxies include the Chicago Fed National Activity Index (CFNAI), the macroeconomic and financial uncertainty indices from Jurado, Ludvigson, and Ng (2015). If bond risk premia are countercyclical, they should have negative correlations with CFNAI, but positive correlations with macro and financial uncertainty.

We report the results in Table B.10. To save space, we only show the results for 2-year and 5-year returns. First, for FB, CP, and CF, almost all correlations between macro variables and bond risk premia inferred from linear regression models are close to 0. For example, linear regression models using CP and CF predictors produce bond risk premia which are positively correlated with CFNAI.

This is against theoretical prediction. Predictive systems produce bond risk premia that are negatvie correlated with CFNAI and positively correlated with macro and financial uncertainty. This indicates that the information in returns is useful in producing such countercyclical dynamics in bond risk premia. Also, when the sign is "correct" (indicating countercyclical risk premia), the absolute values of correlations from predictive systems are generally higher than those from linear regression models. Interestingly, the bond risk premia inferred using uninformative predictive systems produce the highest absolute values of correlations. Next, in the last block for LN predictor, signs of the correlations indicate that bond risk premia are all countercyclical. The correlations inferred from linear regression models have the strongest correlations with macro variables. This is not surprising because the first principal component of macro variables is closely related to real economic activity (Ludvigson and Ng (2009)). In sum, these results show that consistent with the prediction of Wachter (2006), bond risk premia inferred by predictive systems are indeed countercyclical, whereas some risk premia inferred by linear regression models are not.

Next, several papers point out that inflation risk should affect bond returns. Bansal and Shaliastovich (2013) use long-run risk model to show that bond risk premia rise with uncertainty about expected inflation. Creal and Wu (2017) extend the habit model of Campbell and Cochrane (1999) and allow both price and quantity of risk to be time-varying. They also point out that one important underlying risk is inflation risk. Wright (2011) use international evidence to show that inflation risk can be one crucial component in bond risk premia. To empirically test this hypothesis, we use data from Survey of Professional Forecasters published by Philadelphia Fed, and measure inflation uncertainty by cross-sectional forecast dispersion (Wright (2011), Gargano, Pettenuzzo, and Timmermann (2017)). We report the correlations between expected bond returns and inflation uncertainty in the last 2 columns of Table B.10. All of the correlations under predictive systems are significantly positive. When we use linear regression models and CF predictor, bond risk premia are negatively correlated with inflation risk.

In sum, the bond risk premia inferred using predictive systems are countercyclical and increase with inflation risk, and this is consistent with theoretical implications.

3.6 Concluding Remarks

This paper studies bond return predictability in the framework of predictive systems. We take into account a wide range of predictors in the literature. Predictive systems are designed to deal with imperfect predictors and can incorporate prior beliefs about the potentially negative correlation between unexpected and expected returns. We have some interesting findings. First, predictive systems can produce much stronger evidence of predictability than linear regression models. Second, we investigate the link between expected bond returns and a set of macroeconomic variables. We find that bond risk premia inferred using predictive systems are countercyclical and increase with inflation risk. This is in line with the implications of several consumption-based asset pricing models, so predictive systems can produce more economically reasonable bond risk premia, and several macroeconomic variables are potential drivers of its variation.

We highly recommend the use of predictive systems to improve the evidence of equity or bond return predictability. The standard linear regression model may be too restrictive in assuming a perfect linear relationship between expected returns and current value of the predictor, and sometimes it produces counterintuitive bond risk premia.

There are a number of interesting subjects for future research. First, Cochrane and Piazzesi (2005) uncover a common factor from 2-year to 5-year bond excess returns. It would be interesting to identify similar factor structure from different maturity bond returns under predictive systems, and compare it with the factor based on linear regression models. Second, we do not consider stochastic volatility. Some previous studies on equity return predictability (Johannes, Korteweg, and Polson (2014)) have shown benefits of this feature. Third, it may be useful to impose other economically motivated constraints on equity or bond risk premia. For example, Pettenuzzo, Timmermann, and Valkanov (2014) find that ruling out negative equity premium and bounding the conditional Sharpe ratio can improve predictive accuracy. One may think of creating new models (e.g., nonlinear predictive systems) and algorithms to incorporate other features.

Chapter 4

Investor Sentiment and Bond Return Predictability

4.1 Introduction

Equity and Treasury Bond markets are two extremely important targets for today's investors, economists, and policymakers. The bond market is closely linked with monetary policy, and is expected to interact with the stock market. Any tiny change of U.S. interest rate "will likely change the dynamics of investing in the stock market"[1]. Several papers have tried to understand the correlation between stock and government bond returns. For example, Baele, Bekaert, and Inghelbrecht (2010) show that correlation between daily equity and bond returns could vary over time between -0.60 and 0.60, and liquidity proxies play an important role in explaining such correlation. Baker and Wurgler (2012) study the link between government bonds and the cross section of stocks. They observe that stock market sentiment and flights to quality are anecdotally associated over time during special financial market episodes, such as during the recent financial crisis. Their results suggest that sentiment, a predictor of the cross section of stock returns, predicts excess government bond returns. This delivers evidence that the expected returns of stocks and bonds are firmly linked.

However, the above observation of Baker and Wurgler (2012) is drawn on pure in-sample tests. Recently, several studies focus on whether in-sample evidence of bond return predictability could generate economic values for real-time investors. Thornton and Valente (2012) find that using forward spreads as predictor does not

[1]Article by U.S. News Website. Simon Constable: How Rising Interest Rates Will Hurt the Stock Market. https://money.usnews.com/investing/bonds/articles/2017-12-05/how-rising-interest-rates-will-hurt-the-stock-market

create higher utility compared with using expectations hypothesis model, which indicates no predictability. Sarno, Schneider, and Wagner (2016) use affine term structure models and reach similar conclusion. Although Gargano, Pettenuzzo, and Timmermann (2017) find that in-sample evidence is linked with out-of-sample statistical and economic evidence, their tests use fully revised and not real time information. Fulop, Li, and Wan (2018) find that the strong statistical and economic evidence from fully revised macroeconomic data vanishes when real-time and survey-based information is used instead. So, up to now in the literature, in general we do not observe evidence of out-of-sample bond return predictability.

In this paper, we test whether the stock market sentiment index can predict government bond returns both in and out of sample. Traditional tests mainly rely on linear regression model, but we also consider the predictive system model proposed by Pastor and Stambaugh (2009). The predictive system was originally used to estimate equity risk premia and was used recently by Wan (2018) to evaluate bond risk premia. In real time analysis, we avoid hindsight problem and use only information available in real time.

Our results show that we can identify some in-sample evidence of bond return predictability. Using stock market sentiment as predictor, we can generate in-sample R-squared of similar magnitude as when we use information from the yield curve. For example, when we use predictive system model with more informative priors, the R^2 values generated using Fama and Bliss (1987) predictor range from 6.19% to 6.50% for 2-year to 5-year returns, and those generated using sentiment index range from 6.32% to 6.49%. However, when we switch to use only real-time data, both statistical and economic metrics suggest that there is hardly any evidence of predictability. Our additional robustness tests deliver similar observation. Thus, how to prove the link between in-sample and out-of-sample predictability still remains a challenge.

The rest of the paper is organized as follows. Section 4.2 presents variables, predictive models and introduces our methodology. Section 4.3 presents the data and summary statistics. Section 4.4 provides the main empirical results and Section 4.5 concludes the paper.

4.2 Models and Methodology

4.2.1 Linear Regression Model

In the current literature, a lot of papers rely on traditional linear regression models to test bond return predictability. We first define the log-yield of an n-year bond as $y_t^{(n)} \equiv -\frac{1}{n} p_t^{(n)}$, where $p_t^{(n)} = \ln P_t^{(n)}$, and $P_t^{(n)}$ is the nominal price of an n-year zero-coupon Treasury Bond at time t. The excess return of an n-year bond is computed as follows: first we compute the difference between the return from buying an n-year bond at time t and selling it one month later, then we deduct the one-month risk free rate. In this paper, we assume the data frequency is one-month, and maturity of the bond, n, can be 2, 3, 4, or 5 years. Then the linear predictive model is:

$$r_{t+1}^{(n)} = \alpha + \beta X_t + \epsilon_{t+1}, \tag{4.1}$$

where X_t is a set of pre-determined predictors, $\epsilon_t \sim N(0, \sigma_x^2)$ is the error term.

Fama and Bliss (1987) find that the forward spread has predictive power for excess bond returns. Fulop, Li, and Wan (2018) find some real time statistical evidence using information in forward rates. Thus, we use the forward spreads (FB) from Fama and Bliss (1987) as another predictor for comparison. The FB factor is defined as: $FB_t^{(n,1/12)} = f_t^{(n-1/12,n)} - 1/12 \cdot y_t^{(1/12)}$.

The sentiment index of Baker and Wurgler (2006) is based on the common variation in several underlying proxies of sentiment: the closed-end fund discount, the number of IPOs, average first-day returns on IPOs, the equity share in new issues, and the dividend premium.[2] We follow Baker and Wurgler and use the first principal component (PC) of the above proxies to construct the sentiment index. In out-of-sample analysis, we extract the PC using only information available in real time and re-construct the sentiment index.

4.2.2 Predictive System Framework

In addition to linear regression model, we consider a model proposed by Pastor and Stambaugh (2009), the predictive system framework. Pastor and Stambaugh use this framework to estimate equity risk premia: it allows the predictor to be imperfectly correlated with expected returns. Under this framework, the prior beliefs about the correlation between expected and unexpected returns can substantially affect estimates of equity risk premia. Their results show that predictive systems

[2]According to the data published on Jeffrey Wurgler's website, NYSE turnover is no longer considered a proxy of sentiment, given the explosion of high-frequency trading and the migration of trading.

can deliver higher in-sample R-squared when we regress the realized returns on expected returns estimated using predictive system. Recently, Wan (2018) uses this framework in Treasury bond market, and finds that the bond risk premia are more economically reasonable when we use predictive systems instead of linear predictive models.

The predictive system is defined as follows:

$$r_{t+1}^{(n)} = \mu_t + u_{t+1} \tag{4.2}$$

$$X_{t+1} = (1-A)E_X + AX_t + v_{t+1} \tag{4.3}$$

$$\mu_{t+1} = (1-B)E_{rx} + B\mu_t + w_{t+1} \tag{4.4}$$

where μ_t, a latent factor, represents the expected return process. Both the predictor and μ_t follow AR(1) process. The error terms $\{u_t, v_t, w_t\}$ are assumed to be distributed identically and independently across t with a covariance Σ:

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma \equiv \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \sigma_v^2 & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix} \right)$$

Under predictive systems, when the predictor delivers expected returns perfectly, the model will revert back to linear regression model. However, most predictors contain noise. In this case, the expected return process will be a weighted average of both returns and predictors. Also, according to Pastor and Stambaugh (2009), the correlation coefficient ρ_{uw} can substantially affect the estimates of equity risk premia. A negative ρ_{uw} means that unexpected positive shocks to expected returns are likely to be accompanied by unexpected negative realized returns, therefore asset prices decrease when discount rates rise, which is consistent with real-life observation and financial theory. We refer to ρ_{uw} as the "correlation between expected and unexpected returns", which is slightly inaccurate but much simpler. We follow Pastor and Stambaugh (2009) and assume three different priors for ρ_{uw}. More details will be given in Section 4.2.3.

We only consider single-predictor predictive models. In the linear model, when no predictor is used, we turn to the expectations hypothesis (EH) model. The EH model assumes no predictability and serves as a benchmark for comparison with the predictive models.

4.2.3 Estimation Methodology

In full sample analysis, we use the OLS method to estimate parameters from the linear regression models. We use the MCMC sampler to estimate parameters from predictive systems. We follow Pastor and Stambaugh (2009) and assume three different sets of priors on the parameter ρ_{uw}. Similar choice of priors has been adopted by Wan (2018). The more informative prior assumes the prior distribution of ρ_{uw} is largely negative: it has 99.5% of its mass below -0.73 and a mean of -0.90. The less informative prior assumes the prior distribution has nearly 99.5% of its mass below 0 and a mean of -0.65. The non informative prior assumes almost a flat prior distribution of ρ_{uw} between -1 and 1. In this article, whenever we use the terms, "more informative", "less informative", and "non informative", it refers to ρ_{uw}.

In out-of-sample tests, we assume a Bayesian investor who updates her beliefs about parameters as new information accummulates. For the linear model, we employ the particle-based learning algorithm developed by Chopin (2002) and used by Fulop, Li, and Wan (2018). To estimate the parameters in predictive systems, we use the algorithm proposed by Wan (2018), which is a straightforward combination of Gibbs sampling (e.g., Casella and George (1992)) and the learning algorithm of Chopin (2002). In the first step of the algorithm, we use Gibbs sampler to estimate the warming-up part of the data. Then in the out-of-sample period, we update the parameters over time with new information. Whenever we face particle degeneracy problems, we use several Gibbs steps to replenish the particle set.

4.3 Data

Our sample period is from January 1966 to November 2015 (599 months in all). We construct monthly returns on US zero-coupon Treasury bonds with maturity 2-, 3-, 4-, and 5-year using the dataset of Gurkaynak, Sack, and Wright (2007)[3]. We use non-overlapping bond returns and the holding period is one month. Both Bauer and Hamilton (2017) and Gargano, Pettenuzzo, and Timmermann (2017) have discussed the use of non-overlapping returns: this can avoid problems such as the strong serial correlated error terms, and non-overlapping returns can better

[3]The data is available at: http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html.

reflect short-term variation in bond prices.

Our FB factor is also constructed using the Gurkaynak, Sack, and Wright (2007) data. We use the data published on Jeffrey Wurgler's website to construct the updated sentiment index. In Table B.14, we show the data summary statistics. Both mean and standard deviation of bond returns increase with maturity. For example, the mean excess return is about 0.121 percent with a standard deviation of 0.857 for 2-year maturity, whereas it increases to 0.194% with a standard deviation of 1.811 for 5-year maturity. Both skewness and kurtosis decrease with maturity: the skewness and kurtosis for 2-year bond return are 0.493 and 14.956, respectively, and they increase to -0.003 and 6.407 for the 5-year return. The first-order autocorrelations of returns are quite low, ranging from 0.115 to 0.167. The mean and standard deviation for FB factor increase with bond maturity. All skewness for FB factor is below 0, whereas the skewness of sentiment index (ST, henceforth) is positive. The FB and ST factors are highly persistent: the first-order autocorrelation coefficients are all very close to 1, ranging from 0.878 to 0.986. We also computed the correlations between different predictor series. The correlations among FB predictors are very close to 1, ranging from 0.86 to 0.99, whereas correlation between FB and ST ranges from 0.05 to 0.12. This indicates that the information sets contained within the two predictors are quite different.

We also plot the updated sentiment index in Figure C.9. Although in this updated version share turnover is no longer considered as a proxy of sentiment, it experiences very similar variation as the originial index in Baker and Wurgler

(2006). ST shows a spike around 1968-1969, then decreases until around 1975. Then it increases from mid 1970s to mid 1980s. Baker and Wurgler (2006) suggest this may be associated with Reagan-era optimism. Then around 2001, it reaches another spike around the dot-com bubble. Quite surprisingly, sentiment does not capture the recent financial crisis. Similar figure has been shown and discussed by Huang, Jiang, Tu, and Zhou (2015).

4.4 Empirical Results

4.4.1 In-Sample Evidence

Our measure for in-sample evidence of bond return predictability is the adjusted R-squared (R^2). The R^2 for linear regression model is straightforwardly computed by regressing excess returns on the predictor. For the predictive systems, we regress excess returns on the filtered expected returns, μ_t. We report the R^2 values in Table B.15. First, in the upper panel, we show the results first for FB factor. We can see that the predictive systems generate higher R^2 values than linear regression models, and when we impose informative priors on the parameter of interest, ρ_{uw}, we obtain higher R^2 values. The predictive systems with more informative priors generate highest R^2 values in all maturities, ranging from 6.19% to 6.50%, while the linear predictive models generate R^2 values between 1.66% and 2.13%. This suggests that it is useful to combine information from the excess bond return series when we estimate expected returns. Second, when we use sentiment index as predictor, R^2 values from the linear models are slightly smaller than those based on FB factor. But when we use the predictive system framework, the R^2 values from using ST factor are comparable to those based on FB factor. For example, when the model is predictive system with less informative priors, the R^2 values based on FB factor range from 3.73% to 4.44%, whereas the R^2 values from using ST factor range from 3.86% to 4.07%.

To sum up, we can observe some in-sample evidence of bond return predictability when we use the investor sentiment index as predictor, and such evidence of stronger under predictive systems.

4.4.2 Real-Time Evidence

In the out-of-sample analysis, we assume a Bayesian agent who updates her beliefs about economic variables over time. The key requirement is that we need to avoid hindsight problem and can only rely on information available up to each time point. We estimate the predicted expected returns $E(r_{t+1}|I_t)$, where I_t denotes information only available at time t. Our full sample is from January 1966 to November 2015, and we use the first twenty years' data (from January 1966 to December 1985) as training sample, so the second part serves as the out of sample period.

Our statistical measure for real time bond return predictability is the classic out-of-sample R-squared, R^2_{OS}, from Campbell and Thompson (2008):

$$R^2_{OS} = 1 - \frac{SSE_i(T_1, T)}{SSE_{EH}(T_1, T)}, \tag{4.5}$$

where T_1 and T denote the beginning and end of the out-of-sample period, and

SSE_i and SSE_{EH} denote the sum of squared forecast errors from certain predictive model i and the EH benchmark. We use the Clark and West (2007) statistic to evaluate the significance of R^2_{OS}. A positive and significant R^2_{OS} suggests evidence of real time predictability.

We show the R^2_{OS} values in the upper panel of Table B.16. First, only 3 R^2_{OS} values based on FB-Linear and 1 R^2_{OS} value based on FB-Less are positive and weakly significant. All the other values based on FB and ST are negative or close to 0, ranging from -1.44% to 0.78%. Second, although previously we observe using predictive systems can generate higher in-sample R-squared than linear regression models, in real time tests the R^2_{OS} values from the two models are in similar magnitude.

Next, we evaluate whether there exists economic evidence of predictability. We assume a real-time Bayesian investor who updates her investment decisions with respect to new information. She constructs a portfolio with risk-free asset and long term bond, and maximizes her expected utility at each time point in the out of sample. We assume she has a power utility function and a risk aversion of 5. Similar practice has been adopted by Gargano, Pettenuzzo, and Timmermann (2017) and Fulop, Li, and Wan (2018), among others. We follow Campbell and Thompson (2008), and transform the realized utility into certainty equivalent returns (CERs). We use Diebold-Mariano measure to test the significance of CER. A positive and significant CER suggests evidence of real-time economic predictability.

The CER values are shown in the lower panel of Table B.16. We can see that no matter which predictor or model is used, non of the CERs is significant. All CER values are quite close to 0, ranging from -0.80% to 0.65%. Clearly there is a sharp difference between in-sample and real-time evidence of predictability. This is consistent with the findings of a lot of previous papers, such as Thornton and Valente (2012), Fulop, Li, and Wan (2018), etc.

To sum up, although we can identify some in-sample evidence of predictability, we do not observe any statistical or economic evidence when we use only real time information.

4.4.3 Robustness Tests

In this section, we conduct several robustness tests. First, we use different risk aversion coefficients for the Bayesian agent in the asset allocation exercise. Previously we choose 5 as risk aversion coefficient, in this robustness test we choose it to be 3 or 10. We report the computed CERs in Table B.17. We can see that only one CER value is positive and mildly significant, and all other values are insignificant and close to 0, ranging from -0.88% to 0.94%.

Next, we change prior distributions of ρ_{uw} when we use predictive systems. Also, we can modify the prior distribution for coefficients in the linear predictive models according to Fulop, Li, and Wan (2018). However, non of these options signficantly change our results. Thus, our main observation remains the same that there is hardly any out-of-sample evidence of predictability using investor sentiment as predictor.

4.5 Concluding Remarks

The influential paper from Baker and Wurgler (2006, 2007) highlights how equity market investor sentiment affects the cross-section of stock returns. Their follow-up paper (Baker and Wurgler (2012)) suggests that investor sentiment also predicts excess bond returns. However, their results are based on pure in-sample tests. Recently, there is a trend in the literature to test whether previously found in-sample evidence of bond return predictability can turn into real time statistical evidence and economic value (Thornton and Valente (2012), Sarno, Schneider, and Wagner (2016), Gargano, Pettenuzzo, and Timmermann (2017), etc). Thus, we assume a Bayesian investor and updates her beliefs in real time to accommodate for new information. We also consider both the traditional linear regression model and the predictive system framework (Pastor and Stambaugh (2009)) to capture the variation in expected bond returns.

Our results show that although we can identify some in-sample evidence based on sentiment index, in real time exercise this hardly generates statistical evidence or economic value for investors. Consistent with a lot of previous literature (Thornton and Valente (2012), Fulop, Li, and Wan (2018), etc), how to identify real-time statistical and economic predictability still remains a challenge.

There are a few interesting directions for future research. First, we may consider stochastic volatility in the predictive models. Some previous studies (e.g., Johannes, Korteweg, and Polson (2014)) have shown it is worthy of adding this feature in the model. Second, in the original paper by Baker and Wurgler (2006), they point out some regime-switching feature of returns with respect to sentiment index. They show that when beginning-of-period sentiment is low, subsequent returns are high for small stocks, young stocks, high volatility stocks, etc, whereas these stocks earn low subsequent returns when sentiment is high. Thus, it may be useful to introduce regime-switching feature into predictive models when we use the sentiment as predictor. Third, we may consider to use the Partial Least Squares method by Wold (1966) and extended by Kelly and Pruitt (2015) to denoise the sentiment index. The paper by Huang, Jiang, Tu, and Zhou (2015) construct a modified sentiment measure using similar method to predict the aggregate stock return. Last, it may be interesting to use other stock market predictors to capture the variation in bond returns (e.g., Goyal and Welch (2008)). The relationship between the stock and bond markets is extremely important to investors, economists, and policymakers. However, this relationship may be also time-varying. Several papers have tried to explain the decoupling of the two markets, but no consensus exists so far. We expect to see more papers on this topic.

Appendix A

Appendix for Essay Two: Predictive Systems, Real Economy, and Bond Risk Premia

A.1 Prior Distributions for Predictive Systems

Define the predictive system as

$$r_{t+1} = \mu_t + u_{t+1} \tag{A.1}$$

$$X_{t+1} = (1 - A)E_X + AX_t + v_{t+1} \tag{A.2}$$

$$\mu_{t+1} = (1 - B)E_r + B\mu_t + w_{t+1} \tag{A.3}$$

(Note that we define r_t as excess return)

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma \equiv \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \sigma_v^2 & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix}.\right) \tag{A.4}$$

Our choice of priors on (E_x, A, E_r, B) is similar as that of Pastor and Stambaugh (2009). The four priors are independent. We restrict both A and B in between $(-1, 1)$. We choose a slightly informative prior for B, $B \sim N(0.99, 0.15^2)$, to account for the concept that μ will be persistent. A, E_r, and E_x all follow normal distribution with large standard deviations.

We divide the covariance matrix Σ into two parts: $\Sigma_{11} \equiv \begin{bmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{wu} & \sigma_w^2 \end{bmatrix}$, and the rest of elements $\Sigma_{(v)} \equiv (\sigma_v^2, \sigma_{uv}, \sigma_{vw})$. We choose informative prior for Σ_{11} but noninformative one for $\Sigma_{(v)}$. We follow Pastor and Stambaugh (2009) and construct a hypothetical sample in which there are T_2 observations of (u, w) but only $T_1 \ll T_2$ observations of v. We choose T_1 to be 4 and T_2 to be one fifth of the full sample size.

We follow Pastor and Stambaugh (2009) and choose an inverted Wishart prior for Σ_{11}, $\Sigma_{11} \sim IW(T_2\hat{\Sigma}_{11}, T_2 - K)$ and the prior mean is $E(\Sigma_{11}) = \hat{\Sigma}_{11}(T_2/T_2 - 4)$ as we use one predictor. Denote $\hat{\Sigma}_{11} = [M_{11}\ M_{12}; M_{21}\ M_{22}]$. We choose elements of Σ_{11} such that the prior mean of σ_u^2 equals 95% of the sample variance of bond excess returns. The prior mean of σ_w^2 is chosen to set the variance of μ_t equal 5% of the sample variance of bond excess returns, combine with a B of 0.97. To set priors on ρ_{uw}, we assume that M_{12} for the more informative prior case follows a uniform distribution on the interval $(-0.95\sqrt{M_{11}M_{22}}, -0.87\sqrt{M_{11}M_{22}})$; M_{12} for the less informative prior case follows a uniform distribution on the interval $(-0.95\sqrt{M_{11}M_{22}}, -0.35\sqrt{M_{11}M_{22}})$; M_{12} for the noninformative prior case follows a uniform distribution on the interval $(-0.95\sqrt{M_{11}M_{22}}, 0.95\sqrt{M_{11}M_{22}})$. Our choice of more and less informative priors are slightly more aggressive than the choice of Pastor and Stambaugh, to account for the possibility that there may be a more negative correlation between expected and unexpected returns for Treasury bonds.

The priors for elements in $\Sigma_{(v)}$ are noninformative as Pastor and Stambaugh (2009). In particular, we run a regression of v_t on (u_t, w_t) with zero intercept. That is, $C = [\sigma_{uv}\ \ \sigma_{vw}]\Sigma_{11}^{-1}$, and $\Omega = \sigma_v^2 - C\Sigma_{11}C'$. We assume $\Omega \sim IW(T_1\Omega_1, T_1)$ and $\text{vec}(C)|\Omega \sim N(\hat{c}_1, \Omega \otimes (X_1'X_1)^{-1})$. We choose the matrix $X_1'X_1$ to be a small positive number times the identity matrix, such that the prior variance of C is large. Thus, the choices of $\Omega_1, \hat{c}_1$ are inconsequential.

A.2 Bayesian Estimation of Predictive Systems

A.2.1 Full-Sample Estimation

The full sample estimation of predictive systems is based on Gibbs sampling, a Markov Chain Monte Carlo (MCMC) method (Casella and George (1992)). We sample parameters from their prior distritions. Then we draw the parameters (E_r, A, E_x, B, Σ) conditional on current draws of μ_t, then we use the forward filtering, backward sampling algorithm (Carter and Kohn (1994)) to draw the series of μ_t conditional on current draws of (E_r, A, E_x, B, Σ).

A.2.2 Real-Time Estimation

In the out-of-sample estimation, we employ a new algorithm to combine Gibbs sampling and sequential Monte Carlo method by Chopin (2002). In the initial warm-up stage, we use exactly the same options of Gibbs sampling as in full sample estimation. Then the output of parameter distribution will serve as priors for the learning stage, which will greatly speed up the estimation. For each parameter set (E_r, A, E_x, B, Σ), we set equal initial weight and run a Kalman filter as the system is linear. We update the weights of the parameter sets according to the estimated likelihood. If we do not enrich the parameter population, this will lead to a gradual deterioration of the performance of the algorithm. To solve this, whenever the effective sample size is below a fixed threshold, we turn to the resample-move step and use a few Gibbs steps to rejuvenate the parameter population. Note that in the move step, we still alternate between drawing states conditional on parameters and drawing parameters conditional on states.

A.3 Expected Returns and Past Values

Define the predictive system as

$$r_{t+1} = \mu_t + u_{t+1} \tag{A.5}$$

$$X_{t+1} = (1 - A)E_X + AX_t + v_{t+1} \tag{A.6}$$

$$\mu_{t+1} = (1 - B)E_r + B\mu_t + w_{t+1} \tag{A.7}$$

(Note that we define r_t as excess return)

$$\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma \equiv \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \sigma_v^2 & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix}.\right) \tag{A.8}$$

We use similar notations as in the Appendix of Pastor and Stambaugh (2009). By the standard methodology of Kalman filtering, define

$$z_t = [r_t \;\; x_t]',\; E_z = [E_r \;\; E_x]',\; V_{zz} = \begin{bmatrix} V_{rr} & V_{rx} \\ V_{xr} & V_{xx} \end{bmatrix},\; V_{z\mu} = [V_{r\mu} \;\; V_{x\mu}]'$$

$$a_t = E(\mu_t|D_{t-1}),\; b_t = E(\mu_t|D_t),\; e_t = E(z_t|\mu_t, D_{t-1})$$

$$f_t = E(z_t|D_{t-1}),\; P_t = Var(\mu_t|D_{t-1}),\; Q_t = Var(\mu_t|D_t)$$

$R_t = Var(z_t|\mu_t, D_{t-1}),\ S_t = Var(z_t|D_{t-1}),\ G_t = Cov(z_t, \mu_t'|D_{t-1})$

Then we can express the vector of conditional expected returns, $b_t = E(\mu_t|D_t)$, as a function of past returns and predictors. Define

$[M_t\ \ N_t] \equiv P_t(P_t + G_t'R_t^{-1}G_t)^{-1}G_t'R_t^{-1} = G_t'S_t^{-1}$,

then for $t > 1$, $b_t = (1-B)E_r + (B-M)b_{t-1} + M_t r_t + N_t v_t$.

For t=1, $b_1 - E_r = M_1(r_1 - b_0) + N_1 v_1$, where v_1 is innovation in the predictor.

If we repeat substitution, then

$b_t = E_r + \sum_{s=1}^{t} \Lambda_s(r_s - b_{s-1}) + \sum_{s=1}^{t} \Phi_s v_s$, where $\Lambda_s = B^{t-s}M_s$, $\Phi_s = B^{t-s}N_s$.

Thus, the conditional expected return can be seen as a linear combination of past return forecast errors and past innovations in the predictors.

A.4 Out-of-Sample Predictability

A.4.1 Real-Time Data

In out-of-sample tests, we take into account publication delay and data revisions of macroeconomic variables, so at each month t we can only obtain the observation of each macroeconomic variable for month $t-1$ or even month $t-2$. We reconstruct CP, CF, and LN using real-time information at each time. Currently there are two definitions of "real-time" macroeconomic information in the literature, and we define two LN-type predictors accordingly. LNHB ("HB" stands for historical best information) is constructed using the first principal component of the most recently published macroeconomic data. Similar notion of real-time data is adopted by Eriksen (2017) and Giacoletti, Laursen, and Singleton (2018). LNFR ("FR" stands for first releases) is constructed using the first principal component of the first released macroeconomic information, as Ghysels, Horan, and Moench (2018).

A.4.2 Real-Time Learning

In the out-of-sample analysis, we assume a real-time Bayesian investor who faces the same belief updating problem as the econometrician (Hansen (2007)). She needs to learn about parameters, latent states, and models over time when new observation arrives. The information set includes the history of excess bond returns and predictors, $I_t = \{rx_{1:t}^{(n)}, X_{1:t}\}$. $X_{1:t|t}$ denotes the time series of the predictor from time 1 to t based on information only available up to time t and suggests that our predictors are updated in real time. At each time t, Bayesian learning aims to pin down the joint posterior distribution of the unknown parameters, Θ, and the hidden state, μ_t, with observation I_t and model $\mathcal{M}_i$,

$$p(\mu_t, \Theta | I_t, \mathcal{M}_i) = p(\mu_t | \Theta, I_{1:t}, \mathcal{M}_i) p(\Theta | I_t, \mathcal{M}_i), \tag{A.9}$$

where $p(\mu_t | I_t, \Theta, \mathcal{M}_i)$ solves the state filtering problem, and $p(\Theta | I_t, \mathcal{M}_i)$ addresses the parameter learning issue.

For the linear predictive model of Equation (3.2), Bayesian learning is straightforward with the particle-based learning algorithm proposed by Chopin (2002). Similar learning algorithm is used by Fulop, Li, and Wan (2018) in comparing revised and real-time macro predictors. When we estimate the predictive system in real time, we face problems such as multi-dimensional learning and restricted prior beliefs. To take into account both the parameter learning and state filtering over time, we design a new learning algorithm, which takes advantage of both traditional Gibbs sampling (e.g., Casella and George (1992)) and sequential learning algorithm (Chopin (2002)). It easily accounts for the prior distribution but requires much less computational effort than simply repeating Gibbs sampling at each time. In the first/warm-up/in-sample stage of the algorithm, Gibbs sampling is used to pin down the distributions of parameters and states, under different prior beliefs of certain parameters. Then the output (the posterior distribution of $p(\Theta | I_t, \mathcal{M}_i)$) can serve as a prior for the second/learning stage. At each time of the learning period, we use the parallel sequential learning method to update the weight of each parameter set from its distribution. Whenever we face sample degeneracy problem (i.e., the efficient population size is below a certain threshold), we turn to the resample-move step and use several Gibbs steps to replenish the parameter population. In the warm-up stage, we use same priors and hyper-parameters as in Section 3.4.1.

To illustrate the real-time learning process, we first plot some parameter learning results from the predictive system in Figure C.6. In this example, the CP predictor is used to forecast 5-year excess returns. Our out-of-sample learning starts at April 1994. Note that in the Bayesian algorithm, we use the output of Gibbs sampling as prior in the learning stage, so the confidence intervals in 1994 are already quite tight. First, we observe evidence of time-varying parameters. All three variance parameters decrease over time. Variance of unanticipated shocks to the predictor (σ_v^2) decreases from over 0.03 to under 0.01. The AR(1) coefficient of the predictor (A) gradually increases from around 0.82 to above 0.9, whereas B, E_{rx}, and E_x do not show much variation. Second, some parameters are easier to learn than the others. The confidence intervals for σ_v^2 and A converge as more information accumulates, whereas the confidence intervals for other parameters do not tighten up over time.

In Figure C.7 we plot the sequential learning results for the key parameter, ρ_{uw}. Confidence interval from the noninformative prior is the largest among the three panels, and it shrinks when we switch to more aggressive priors. The mean of ρ_{uw} from the noninformative prior is negative at the beginning and increases to nearly 0.5 then decreases. The distributions of ρ_{uw} from the more and the less informative priors show very slight variation over time. Different priors will certainly impact the estimates of conditional expected returns (μ_t). As an example, in Figure C.8 we plot the real-time forecasts of conditional expected returns from different specifications. The estimates from the linear regression model seem extremely volatile, and estimates from predictive systems are much smoother and more stable over time. The more informative prior provides the most persistent estimates of expected returns, and the degree of smoothing decreases as we use less aggressive priors.

Thanks to the learning algorithm, we obtain the predictive distribution of the bond returns at each time,

$$p(rx_{t+1}^{(n)}|I_t, \mathcal{M}_i) = \int p(rx_{t+1}^{(n)}|\mu_t, \Theta, I_t, \mathcal{M}_i)p(\mu_t|\Theta, I_t, \mathcal{M}_i)p(\Theta|I_t, \mathcal{M}_i)d\mu_t d\Theta. \tag{A.10}$$

We will use this distribution to compare statistical and economic performances of different models.

A.4.3 Statistical Evidence

In the out-of-sample analysis, given the full predictive distribution of excess returns, we can use the posterior mean as the point forecast of each model at each time t. Campbell and Thompson (2008) define the statistical performance measure, the out-of-sample R^2 (R^2_{OS}), as

$$R^2_{OS} = 1 - \frac{SSE_i(T_0, T)}{SSE_{EH}(T_0, T)}, \tag{A.11}$$

where T_0 and T denote the beginning and end of the out-of-sample period, and SSE_i and SSE_{EH} denote the sum of squared forecast errors from the predictive model and the EH benchmark. The R^2_{OS} is analogous to the standard R^2 and a positive R^2_{OS} suggests evidence of return predictability. We use the statistic developed by Clark and West (2007) to evaluate the significance of R^2_{OS}.

Our full sample is from April 1982 to December 2015, and we choose the first 12 years to be the warm-up stage for the learning algorithm, so out-of-sample period starts in April 1994. We report the statistical evidence of predictability in Panel A of Table B.11. First, for the linear regression models, FB and CF can generate positive and significant R^2_{OS}, and the evidence for the 5-year return is only weakly significant. R^2_{OS}'s for CP and LN predictors are all negative. Second, the more complex predictive systems generally outperform the simpler linear regression models, except for CF with long maturity returns. R^2_{OS}'s from predictive systems in forecasting 2-year returns are all significant, but their values and level of significance gradually decrease as maturity increases. This pattern is similar to what we find from the in-sample analysis. R^2_{OS}'s from FB and CP using predictive systems are all positive. This confirms the advantage of using predictive systems, and if we use simple linear regression models the predictive ability of predictors such as CP would have been ignored. Third, comparing within the systems, the noninformative predictive system produces the highest R^2_{OS} in the short maturity, but its performance deteriorates with maturity and is gradually caught up by the more and less informative systems. Last, of all predictors, FB produces the highest

R^2_{OS} and those values are significant for different maturities. This again highlights the importance of information from the term structure in explaining bond return variations.

In sum, the results from Panel A prove some statistical evidence of out-of-sample predictability and underline the importance of using predictive systems. However, the real-time evidence is weaker than in full-sample tests. This is consistent with the findings of Cieslak and Povala (2015), Haddad and Sraer (2018), and Farmer, Schmidt, and Timmermann (2018). Next we investigate whether such statistical evidence can translate into economic gains for the Bayesian investor.

A.4.4 Economic Evidence

To evaluate the out-of-sample economic predictability, we consider the asset allocation decisions of a real-time Bayesian investor who selects optimal weights for a risk-free one-month interest rate and a risky bond with maturity n, to maximize her expected utility of the next period portfolio value. Following the practices of Gargano, Pettenuzzo, and Timmermann (2017) and Fulop, Li, and Wan (2018), we assume a power-utility and transform the investor's realized utility into certainty equivalent returns (CERs). A positive and significant CER indicates evidence of economic predictability.

Panel B of Table B.11 presents CERs from different specifications. We set the risk aversion to be 5 and restrict the investment weights on risk-free rate and long-term bond yield in between -1 and 2 to avoid extreme investments (Gargano, Pettenuzzo, and Timmermann (2017)). We find that no matter which predictor is used and which maturity is considered, non of the CERs is significant. All the values are either negative or close to 0. Predictive systems using FB, CP, and LNFR on average generate higher CERs than the corresponding linear regression model, whereas systems using CF underform.

The above results clearly indicate a gap between statistical and economic metrics of predictive performance. Similar gap is also identified by Goyal and Welch (2008), Thornton and Valente (2012) and Sarno, Schneider, and Wagner (2016).

In addition, in some previous out-of-sample results, we observe that B (the AR(1) coefficient of the expected returns) can drop to below 0.9. Ferson, Sarkissian, and Simin (2003) discuss several reasons to believe μ_t for equity returns is persistent, and we think the logic could be analogous for bond returns. Therefore we force the AR(1) coefficient of μ_t, B, to be 0.95 in the predictive systems, and test whether this will improve the results. In another setting, we assume a higher risk aversion coefficient ($\gamma = 10$) for the Bayesian investor. However, unreported results from these tests do not significantly improve the out-of-sample performance.

A.4.5 Link to Real Economy: Real-Time Analysis

In this section, we study how real-time bond excess return forecasts relate to macroeconomic variables. First, we report the correlations between out-of-sample expected return forecasts and macroeconomic condition proxies in Table B.12. For FB, CP, and CF predictors, we find many more procyclical risk premia from both linear regression model and predictive systems than in the full-sample analysis. For example, using FB and more informative predictive system now produces positive correlation with CFNAI and negative correlation with macro uncertainty. For the two real-time LN predictors, linear regression models still produce significantly countercyclical risk premia. This is not surprising because the first principal component is closely linked to real growth. But using LN and predictive systems can produce some procyclical risk premia (e.g., predictive system with more informative prior). One might think that the difference between full-sample and real-time results are due to the choice of sample period (as the out-of-sample period is the latter part of full-sample), but when we test correlations between macro conditions and expected returns during the same period (from April 1994 to December 2015), we still observe similar patterns. Second, we compute the correlations between out-of-sample expected return forecasts and inflation uncertainty. We find more negative values than in full-sample tests, and those positive values are generally smaller and less significant.

In sum, the above results show that real-time interactions between bond risk premia and macro variables are weaker. Bond risk premia estimated using full sample information may be more accurate.

A.4.6 Model Combinations

In addition to parameter uncertainty and state uncertainty, the Bayesian investor faces model uncertainty. Gargano, Pettenuzzo, and Timmermann (2017) argue that model combination can stabilize forecasts relative to those from individual models, and work similarly as a diversification strategy to improve portfolio performance. In this section, we introduce different schemes to combine the predictive power of individual models, including predictive regressions and predictive systems.

We use the same combination schemes as Fulop, Li, and Wan (2018): Sequential Best Model (SBM) chooses the single model with the largest marginal likelihood at each time t; Bayesian Model Averaging (BMA) distributes weights according to the marginal likelihood of each model; Equal-weighted Model Averaging (EMA) assumes equal weight on each model; Utility-based Model Averaing (UMA) distributes weights according to realized utility.

Table B.13 presents the statistical and economic results for the combination schemes. We compare each scheme with the no-predictability EH model. In Panel A, R^2_{OS}'s are all positive and significant, but the magnitude of R^2_{OS} and level of significance decrease with return maturity. Panels B presents the CERs. We test the economic performance using a risk aversion of 5. Although we observe some positive CERs, they are all insignificant and close to 0.

Appendix B

Tables

Table B.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 for parameters in different models. The linear model is given in Equation (3.2). Parameters for the linear models are: α, β, and σ_{rx} The stochastic volatility model is given in Equations (2.5) and (2.6). Parameters for the SV models are: α, β, μ, ϕ, and σ_h.

Table B.2: **Full-Sample Summary Statistics**

Panel A: Excess 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	LN	LNRT1	LNRT2	LNRT3	LNSF
Mean	0.104	0.128	0.147	0.162	0.146	0.146	0.146	0.146	0.270	0.428
St.dev	0.095	0.110	0.122	0.132	0.197	0.265	0.107	0.092	0.060	0.535
Skew	0.015	-0.233	-0.265	-0.212	0.668	0.402	-1.372	1.254	1.042	-1.622
Kurt	4.016	3.642	3.283	2.965	4.466	3.469	9.876	6.639	5.935	6.398
AC(1)	0.880	0.899	0.913	0.923	0.703	0.471	0.542	0.718	0.711	0.896

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. Full-sample data is from January 1962 to September 2017. For LNRT3, the full sample is from April 1982 to December 2015, and for LNSF, the full sample is from 1968:Q4 to 2014:Q4.

Table B.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
FB-SV	−2.17	0.25	1.86*	2.81**	−1.03	−0.89	−0.04	0.87
CP-CV	0.39*	0.59*	0.48	1.42	−0.68	−0.84	−0.59	0.00
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
FB-SV	−1.07	−1.32	−1.17	−0.36	−0.35	0.35	0.70	0.87
CP-CV	−0.69	−1.15	−1.74	−1.14	−0.12	0.11	0.07	0.24
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-squared, R^2_{OS}, and annualized CERs (in percentage) 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 B.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
LN-CV	−13.40	−4.44	−0.61	0.64*	0.24	0.91	2.19***	2.90***
LN-SV	3.08**	3.74**	3.83**	3.62**	0.24	0.89	2.01**	2.55**
LNRT1-CV	−38.04	−25.25	−17.35	−12.03	−0.38	−0.47	0.06	0.17
LNRT1-SV	−15.11	−22.45	−19.05	−10.28	−0.66	−2.04	−1.94	−1.22
LNRT2-CV	0.13	−0.44	−0.11	0.32	−0.03	0.09	0.33	0.71
LNRT2-SV	−0.28	−1.32	−1.60	−1.20	−0.13	−0.16	0.00	−0.17
LNRT3-CV	−4.86	−3.35	−2.69	−1.80	0.01	0.00	−0.09	−0.18
LNRT3-SV	−0.30	−2.64	−2.27	−1.31	0.01	0.02	−0.08	−0.15
LNSF-CV	−13.18	−7.86	−5.66	−2.93	0.16	0.53	0.64	0.93
LNSF-SV	−9.64	−6.62	−3.16	−2.29	0.10	0.40	0.76	0.89
	B. CER: $\gamma = 3$				D. CER: $\gamma = 10$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
LN-CV	0.20	0.42	0.82	1.54	0.92**	2.02***	2.40***	2.15*
LN-SV	0.15	0.53	0.68	1.43	0.89**	1.97***	2.29**	1.93*
LNRT1-CV	−0.25	−0.63	−0.72	−0.33	−0.01	−0.05	−0.03	−0.01
LNRT1-SV	−0.51	−2.18	−3.26	−2.71	−0.45	−0.70	−0.75	−0.45
LNRT2-CV	0.00	−0.08	0.06	0.63	0.47**	0.31	0.17	0.20
LNRT2-SV	−0.04	−0.37	−0.97	−0.67	0.40	0.77	0.14	−0.24
LNRT3-CV	0.00	0.02	0.01	0.02	−0.01	−0.15	−0.07	−0.23
LNRT3-SV	0.00	0.01	0.01	0.03	−0.01	−0.17	0.01	−0.22
LNSF-CV	0.00	0.16	0.53	0.73	0.39	0.59	0.55	0.61
LNSF-SV	0.00	0.00	0.34	0.61	0.54	0.58	0.77	0.65

This table presents the out-of-sample R-squared, R^2_{OS}, and annualized CERs (in percentage) for the predictive models based on macro factors, LN, LNRT1, LNRT2, LNRT3 and LNSF. 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. For LN and LNRT1/LNRT2, the out-of-sample period is from August 1999 to September 2017. For LNRT3, the out-of-sample period is from April 1994 to December 2015; and for LNSF, the out-of-sample period is from 1994:Q1 to 2014:Q4.

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

Panel A: FB, CP, and LN								
	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 LNRT1								
	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 LNRT2								
	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

(Continued)

Panel D: FB, CP, and LNRT3

	R^2_{OS}				CER: $\gamma = 5$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	10.69***	4.07**	2.33**	2.71***	0.06	−0.19	−0.14	−0.11
BMA	−5.49	−3.19	−2.22	−0.97	−1.96	−2.61	−2.82	−2.64
EMA	3.82***	2.08**	1.44**	1.42**	−0.03	−0.16	−0.40	−0.69
UMA	−0.52	−2.43	0.72*	0.64	−0.07	−0.23	−0.67	−1.03
	CER: $\gamma = 3$				CER: $\gamma = 10$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	0.06	−0.21	−0.39	−0.50	0.03	−0.21	−0.04	0.08
BMA	−2.03	−2.55	−2.88	−2.71	−1.90	−2.12	−1.70	−1.20
EMA	0.03	0.02	−0.08	−0.20	−0.18	−0.51	−0.28	−0.12
UMA	−0.05	−0.16	−0.37	−0.46	−0.17	−0.37	−0.53	−0.38

Panel E: FB, CP, and LNSF

	R^2_{OS}				CER: $\gamma = 5$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	−1.37	−6.69	−2.51	−0.52	−0.04	−0.07	1.05	1.24
BMA	−3.43	−4.81	−1.43	−1.58	−0.74	−0.59	0.00	−0.07
EMA	1.46	2.31	3.49*	3.77*	−0.29	0.15	0.67	0.98*
UMA	−14.57	−10.39	−5.80	−1.38	−0.23	−0.25	0.01	0.46
	CER: $\gamma = 3$				CER: $\gamma = 10$			
	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
SBM	−0.24	−0.86	0.19	0.99	0.52	0.35	0.83	0.84
BMA	−0.96	−1.69	−0.93	−0.72	0.05	0.13	0.36	0.29
EMA	−0.36	−0.49	−0.15	0.34	0.29	0.56*	0.66*	0.68*
UMA	−0.40	−1.03	−0.86	−0.14	0.29	0.31	0.25	0.37

This table presents the out-of-sample R-squared, R^2_{OS}, and annualized CERs (%) for model combinations. The portfolio weight is between -1 and 2. * denotes significance at 10% level, ** significance at 5% level, and *** significance at 1% level. For panels A, B and C, out-of-sample period is from August 1999 to September 2017. For panel D, out-of-sample period is from April 1994 to December 2015. For panel E, out-of-sample period is from 1994:Q1 to 2014:Q4.

Table B.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
LN-CV	7.03**	4.90	2.80	1.19
LNRT1-CV	0.68	0.45	0.37	0.33
LNRT2-CV	1.51**	0.58	0.34	0.41
LNRT3-CV	−0.10	−0.92	−1.34	−1.14
LNSF-CV	1.03*	1.14	1.06	1.17
FB-SV	0.28	0.75	0.85	1.21
CP-SV	2.26	−0.14	−0.04	0.54
LN-SV	11.03**	6.67	2.87	2.00
LNRT1-SV	1.95	−0.17	−0.77	−0.22
LNRT2-SV	4.20	0.33	−0.99	−0.92
LNRT3-SV	0.06	−1.36	−1.08	−0.95
LNSF-SV	2.31*	1.40	1.51	1.33

This table presents annualized CERs (in percentage) 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. For FB, CP, LN, LNRT1 and LNRT2, the out-of-sample period is from August 1999 to September 2017. For LNRT3, the out-of-sample period is from April 1994 to December 2015. For LNSF, the out-of-sample period is from 1994:Q1 to 2014:Q4.

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

	2-year	3-year	4-year	5-year
Panel A: FB, CP, and LN				
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 LNRT1				
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 LNRT2				
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
Panel D: FB, CP, and LNRT3				
SBM	2.65	−0.65	−0.60	0.09
BMA	−6.23	−4.66	−3.43	−2.43
EMA	−1.26	−0.59	−0.54	−0.25
UMA	−0.70	−1.63	−1.02	−0.77
Panel E: FB, CP, and LNSF				
SBM	1.68	0.60	1.34	1.44
BMA	1.93	0.35	0.65	0.48
EMA	0.91*	1.06*	1.25*	1.29*
UMA	1.07	0.59	0.40	0.70

This table presents annualized CERs (in percentage) 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. For panels A, B and C, the out-of-sample period is from August 1999 to September 2017. For panel D, the out-of-sample period is from April 1994 to December 2015. For panel E, the out-of-sample data is quarterly from 1994:Q1 to 2014:Q4.

Table B.8: **Summary Statistics**

Panel A: Excess Bond Returns

	2-Year	3-Year	4-Year	5-Year
Mean	0.19	0.25	0.31	0.36
St.dev	0.63	0.97	1.29	1.61
Skew	0.36	0.12	0.04	0.03
Kurt	4.61	3.83	3.56	3.52
AC(1)	0.22	0.17	0.14	0.12

Panel B: Predictors

	FB2	FB3	FB4	FB5	CP	CF	LN
Mean	0.12	0.16	0.19	0.21	0.27	0.27	0.27
St.dev	0.09	0.10	0.11	0.12	0.20	0.12	0.17
Skew	0.45	0.06	-0.12	-0.18	0.37	-0.08	1.36
Kurt	3.66	2.80	2.41	2.23	2.65	2.28	7.34
AC(1)	0.90	0.92	0.93	0.93	0.94	0.94	0.81

Panel C: Correlation Matrix

	FB2	FB3	FB4	FB5	CP	CF	LN
FB2	1.00	0.96	0.88	0.81	0.62	0.57	-0.11
FB3		1.00	0.98	0.94	0.59	0.60	-0.01
FB4			1.00	0.99	0.53	0.59	0.06
FB5				1.00	0.45	0.57	0.13
CP					1.00	0.37	-0.10
CF						1.00	-0.02
LN							1.00

This table presents the summary statistics of bond excess returns and full-sample predictors. Panel A and B report the mean, standard deviation, skewness, kurtosis, and first-order autocorrelation of monthly excess returns and predictors (in percentage). Panel C presents the correlation matrix of predictors. Sample period is from April 1982 to December 2015.

Table B.9: **Explanatory Power of Predictive Models and Correlation Matrix of Expected Returns**

	Panel A: Full-Sample R^2				Panel B: Correlations of Expected Returns				
	2-Yr	3-Yr	4-Yr	5-Yr		Linear	More	Less	Non
FB-Linear	4.12	2.83	2.26	1.93	FB-Linear	1.00	0.07	0.50	0.55
FB-More	7.97	6.01	3.68	3.37	FB-More		1.00	0.83	0.30
FB-Less	7.46	5.23	4.20	3.71	FB-Less			1.00	0.70
FB-Non	6.65	5.04	4.14	3.74	FB-Non				1.00
CP-Linear	4.34	3.36	3.00	2.81	CP-Linear	1.00	0.48	0.69	0.75
CP-More	7.80	5.85	3.65	3.54	CP-More		1.00	0.93	0.38
CP-Less	6.36	4.66	4.20	3.57	CP-Less			1.00	0.62
CP-Non	6.09	4.22	3.71	3.39	CP-Non				1.00
CF-Linear	0.79	1.03	1.23	1.39	CF-Linear	1.00	0.23	0.42	0.33
CF-More	8.69	7.05	6.36	5.89	CF-More		1.00	0.86	0.58
CF-Less	8.99	7.31	6.43	5.81	CF-Less			1.00	0.90
CF-Non	7.38	6.04	5.34	4.90	CF-Non				1.00
LN-Linear	3.72	2.79	2.13	1.66	LN-Linear	1.00	0.07	0.12	0.36
LN-More	7.96	6.20	4.95	4.36	LN-More		1.00	0.96	0.22
LN-Less	5.97	4.44	3.71	3.19	LN-Less			1.00	0.35
LN-Non	5.27	3.52	2.92	2.71	LN-Non				1.00

Panel A presents the full-sample adjusted R^2 (in percentage) for FB, CP, CF and LN predictors under different predictive models. Within each block, "More", "Less", and "Non" represent results from the predictive systems with three different prior beliefs on ρ_{uw}. Panel B presents the correlation matrix of expected returns for the same predictor under different models to predict 3-year excess returns. Full-sample period is from April 1982 to December 2015.

Table B.10: **Correlations with Macroeconomic Variables**

	CFNAI		Macro Uncertainty		Fin Uncertainty		Inflation Uncertainty	
Model	2-Yr	5-Yr	2-Yr	5-Yr	2-Yr	5-Yr	2-Yr	5-Yr
FB-Linear	0.16***	−0.03	−0.01	0.10**	0.03	0.15***	0.15*	0.20**
FB-More	−0.05	−0.12**	0.21***	0.25***	0.08	0.12**	0.26***	0.23***
FB-Less	−0.11**	−0.21***	0.25***	0.30***	0.18***	0.24***	0.32***	0.28***
FB-Non	−0.15***	−0.34***	0.31***	0.41***	0.22***	0.33***	0.36***	0.38***
CP-Linear	0.09*	0.09*	0.08*	0.08*	0.02	0.02	0.32***	0.32***
CP-More	−0.06	−0.06	0.21***	0.21***	0.08	0.08	0.26***	0.23***
CP-Less	−0.12**	−0.07	0.25***	0.20***	0.16***	0.10**	0.31***	0.27***
CP-Non	−0.19***	−0.24***	0.33***	0.35***	0.24***	0.23***	0.37***	0.37***
CF-Linear	0.12**	0.12**	−0.24***	−0.24***	0.06	0.06	−0.21***	−0.21***
CF-More	−0.07	−0.13**	0.20***	0.23***	0.12***	0.19***	0.23***	0.21**
CF-Less	−0.15***	−0.22***	0.25***	0.25***	0.24***	0.31***	0.26***	0.21**
CF-Non	−0.19***	−0.27***	0.31***	0.31***	0.28***	0.34***	0.32***	0.26***
LN-Linear	−0.89***	−0.89***	0.68***	0.68***	0.50***	0.50***	0.38***	0.38***
LN-More	−0.07	−0.11**	0.24***	0.24***	0.09*	0.10*	0.26***	0.22***
LN-Less	−0.14***	−0.10**	0.27***	0.18*	0.17***	0.11**	0.30***	0.19**
LN-Non	−0.23***	−0.34***	0.34***	0.37***	0.27***	0.24***	0.36***	0.30***

This table presents correlations between macroeconomic variables and expected bond returns. We report results for 2-year and 5-year maturity returns. CFNAI is the Chicago Fed National Activity Index. Macro and financial uncertainty are from Jurado, Ludvigson, and Ng (2015). Inflation uncertainty is from Survey of Professional Forecasters. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. The full-sample period is from April 1982 to December 2015.

Table B.11: **Out-of-Sample Predictability: R^2_{OS} and CER**

	Panel A: R^2_{OS}				Panel B: CER: $\gamma = 5$			
Model	2-Yr	3-Yr	4-Yr	5-Yr	2-Yr	3-Yr	4-Yr	5-Yr
FB-Linear	1.41***	1.66**	1.73**	1.13*	−0.75	−0.89	−0.87	−0.96
FB-More	2.96***	1.65**	1.85**	1.87***	−0.05	−0.33	−0.12	−0.03
FB-Less	4.00***	2.04**	1.28**	1.27**	0.01	−0.05	0.02	−0.11
FB-Non	4.96***	1.78**	0.80	0.66	−0.01	−0.06	−0.09	−0.04
CP-Linear	−1.34	−2.33	−2.66	−2.12	−1.77	−2.48	−2.85	−2.91
CP-More	2.42***	0.51**	0.72	0.89	−0.31	−1.01	−0.80	−0.79
CP-Less	3.76***	1.56**	0.33	0.58	−0.04	−0.19	−0.72	−0.89
CP-Non	4.23***	1.65**	0.78*	0.03	−0.08	−0.20	−0.43	−0.89
CF-Linear	2.56**	1.58**	0.59*	0.55*	−0.04	−0.44	−0.54	−0.34
CF-More	1.20***	−1.38	−1.97	−2.33	−0.70	−1.02	−1.37	−1.67
CF-Less	3.04***	0.32**	−1.18	−2.17	−0.52	−1.03	−1.53	−1.95
CF-Non	3.31***	0.34**	−1.19	−2.16	−0.22	−0.87	−1.59	−2.14
LNHB-Linear	−4.84	−3.30	−2.54	−1.89	0.01	0.01	−0.06	−0.24
LNHB-More	2.77**	0.82	0.14	0.16	0.01	−0.04	−0.21	−0.26
LNHB-Less	2.65***	0.87	−0.23	−0.30	0.00	0.03	−0.01	−0.07
LNHB-Non	3.55***	0.82	−0.29	−0.65	0.00	0.00	−0.01	−0.09
LNFR-Linear	−7.26	−5.02	−4.22	−3.35	0.02	0.01	−0.18	−0.32
LNFR-More	3.11***	0.76	0.50	0.30	−0.01	0.01	−0.05	−0.05
LNFR-Less	2.95***	0.92**	−0.21	−0.37	0.00	0.00	−0.01	−0.01
LNFR-Non	3.10***	0.33	−0.19	0.24	0.00	−0.01	−0.02	−0.10

This table presents R^2_{OS}, and annualized CERs (in percentage) for linear regression models and predictive systems. For each predictor, "More", "Less", and "Non" represent results from the predictive systems using the more informative, the less informative and the noninformative priors on ρ_{uw}. When computing CERs, the risk aversion is 5 and 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 CER is measured using the one-sided Diebold-Mariano statistic. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. The out-of-sample period is from April 1994 to December 2015.

Table B.12: **Out-of-Sample Correlations with Macroeconomic Variables**

	CFNAI		Macro Uncertainty		Fin Uncertainty		Inflation Uncertainty	
Model	2-Yr	5-Yr	2-Yr	5-Yr	2-Yr	5-Yr	2-Yr	5-Yr
FB-Linear	0.01	−0.22***	0.05	0.25***	0.08	0.26***	0.16	0.37***
FB-More	0.13**	0.16*	−0.05	−0.11*	−0.06	−0.17***	0.17	0.10
FB-Less	−0.08	−0.03	0.06	−0.01	0.31***	0.09	0.09	0.16
FB-Non	−0.30***	−0.36***	0.21***	0.23***	0.48***	0.42***	0.08	0.20*
CP-Linear	−0.05	−0.05	0.01	0.03	0.02	0.01	0.19*	0.23**
CP-More	−0.26***	−0.03	0.29***	−0.05	0.11*	−0.12*	0.45***	0.06
CP-Less	−0.31***	−0.20***	0.18***	0.06	0.32***	0.17***	0.22**	0.19*
CP-Non	−0.45***	−0.44***	0.36***	0.36***	0.50***	0.38***	0.27***	0.38***
CF-Linear	0.28***	0.28***	−0.35***	−0.35***	0.04	0.03	−0.32***	−0.31***
CF-More	0.09	0.14*	−0.26***	−0.31***	0.22***	0.16***	−0.24**	−0.25**
CF-Less	−0.13**	−0.03	−0.02	−0.14**	0.41***	0.32***	−0.08	−0.13
CF-Non	−0.23***	−0.07	0.08	−0.10	0.48***	0.35***	0.00	−0.10
LNHB-Linear	−0.55***	−0.56***	0.59***	0.58***	0.51***	0.49***	0.29***	0.30***
LNHB-More	0.46***	0.56***	−0.64***	−0.66***	−0.29***	−0.49***	−0.53***	−0.57***
LNHB-Less	0.21***	0.16***	−0.42***	−0.33***	0.01	−0.27***	−0.37***	−0.33***
LNHB-Non	−0.29***	−0.33***	0.13**	0.18***	0.43***	0.25***	−0.02	0.04
LNFR-Linear	−0.58***	−0.62***	0.65***	0.68***	0.55***	0.55***	0.35***	0.39***
LNFR-More	0.47***	0.42***	−0.63***	−0.48***	−0.23***	−0.36***	−0.46***	−0.51***
LNFR-Less	−0.16***	−0.19***	0.08	0.19***	0.36***	0.20***	−0.01	0.03
LNFR-Non	−0.42***	−0.19***	0.34***	0.07	0.55***	0.25***	0.12	−0.10

This table presents the out-of-sample correlations between macroeconomic conditions and expected bond returns. We report results for 2-year and 5-year maturity returns. CFNAI is the Chicago Fed National Activity Index. Macro and financial uncertainty are from Jurado, Ludvigson, and Ng (2015). Inflation uncertainty is from Survey of Professional Forecasters. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. The out-of-sample period is from April 1994 to December 2015.

Table B.13: **Out-of-Sample Predictability: Model Combinations**

	2-Yr	3-Yr	4-Yr	5-Yr
Panel A: R^2_{OS}				
SBM	5.67∗∗∗	2.15∗∗∗	0.65∗∗	1.55∗∗
BMA	4.53∗∗∗	1.41∗∗	0.72∗	0.90∗
EMA	4.48∗∗∗	2.29∗∗	1.39∗∗	1.37∗∗
UMA	4.53∗∗∗	2.01∗∗	1.63∗∗	1.07∗
Panel B: CER ($\gamma = 5$)				
SBM	0.22	0.21	−0.21	0.06
BMA	−0.05	−0.15	−0.71	−0.99
EMA	−0.01	0.01	−0.03	−0.21
UMA	−0.68	0.04	−0.47	−0.27

This table presents R^2_{OS}, and annualized CERs (in percentage) for model combinations. In panel B, the portfolio weight is restricted in between -1 and 2, and we choose risk aversion to be 5. The statistical significance of R^2_{OS} is measured using the Clark and West (2007) statistic, and the statistical significance of CERs is measured using the one-sided Diebold-Mariano statistic. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. The out-of-sample period is from April 1994 to December 2015.

Table B.14: **Full-sample Summary Statistics**

Panel A: Excess Bond Returns				
	2-Year	3-Year	4-Year	5-Year
Mean	0.121	0.150	0.174	0.194
St.dev	0.857	1.201	1.513	1.811
Skew	0.493	0.194	0.040	-0.003
Kurt	14.956	10.613	7.780	6.407
AC(1)	0.167	0.149	0.131	0.115

Panel B: Predictors					
	FB2	FB3	FB4	FB5	Sentiment
Mean	0.109	0.133	0.153	0.169	0.000
St.dev	0.099	0.115	0.127	0.137	1.000
Skew	-0.111	-0.350	-0.385	-0.337	0.398
Kurt	3.802	3.502	3.191	2.897	3.500
AC(1)	0.878	0.897	0.910	0.920	0.986

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. Full-sample data is from January 1966 to November 2015.

Table B.15: **In-Sample Predictability:** R^2

FB	2-Yr	3-Yr	4-Yr	5-Yr
More informative	6.50	6.35	6.26	6.19
Less informative	4.44	4.25	3.73	3.95
Non informative	3.20	3.12	3.22	3.42
Linear model	1.75	1.66	1.88	2.13
ST	**2-Yr**	**3-Yr**	**4-Yr**	**5-Yr**
More informative	6.49	6.37	6.35	6.32
Less informative	4.07	4.04	3.86	3.90
Non informative	2.86	2.88	2.65	2.70
Linear model	0.49	0.39	0.33	0.29

This table presents the in-sample R-squared, R^2 (in percentage) for FB and ST predictors. More, less, and non informative specifications show results from predictive systems using different priors on ρ_{uw}. The full sample period is from January 1966 to November 2015.

Table B.16: **Out-of-Sample Predictability: R^2_{OS} and CER**

Panel A: R^2_{OS} (%)				
	2-Yr	3-Yr	4-Yr	5-Yr
FB-Linear	-0.42	1.28**	0.87*	1.44**
FB-More informative	-0.17	0.27	-0.38	-0.23
FB-Less informative	-1.44	0.50*	-0.11	-0.17
FB-Non informative	0.26	0.27	0.07	-0.22
ST-Linear	0.31	0.59	0.01	0.09
ST-More informative	0.18	0.25	-0.23	0.12
ST-Less informative	-0.37	-0.01	-0.10	-0.27
ST-Non informative	0.78	0.62	-0.12	-0.18
Panel B: CER (%)				
	2-Yr	3-Yr	4-Yr	5-Yr
FB-Linear	-0.80	-0.23	0.10	0.65
FB-More informative	0.01	0.05	-0.32	-0.26
FB-Less informative	-0.19	0.15	-0.12	-0.19
FB-Non informative	0.03	0.12	-0.02	-0.20
ST-Linear	-0.13	0.07	-0.01	0.09
ST-More informative	-0.03	0.06	-0.21	-0.03
ST-Less informative	0.05	-0.05	-0.04	-0.26
ST-Non informative	0.06	0.22	-0.16	-0.22

This table presents the out-of-sample R-squared, R^2_{OS}, and annualized CERs (in percentage) for the predictive models based on FB and ST. The portfolio weight is restricted in between -1 and 2. Risk aversion is 5. 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 January 1986 to November 2015.

Table B.17: **Robustness Tests: Different Risk Aversion**

Panel A: CER (Risk Aversion=3) (%)				
	2-Yr	3-Yr	4-Yr	5-Yr
FB-Linear	-0.88	-0.79	-0.73	-0.09
FB-More informative	-0.01	0.07	-0.31	-0.47
FB-Less informative	-0.17	0.30**	-0.02	-0.38
FB-Non informative	-0.01	0.20	-0.01	-0.28
ST-Linear	-0.01	-0.07	-0.21	-0.11
ST-More informative	-0.01	0.09	-0.24	-0.16
ST-Less informative	-0.01	0.09	0.16	-0.44
ST-Non informative	-0.02	0.17	-0.15	-0.40
Panel B: CER (Risk Aversion=10) (%)				
	2-Yr	3-Yr	4-Yr	5-Yr
FB-Linear	-0.10	0.60	0.70	0.94
FB-More	0.02	0.06	-0.16	-0.13
FB-Less	-0.22	0.10	-0.06	-0.09
FB-Non	0.04	0.04	-0.01	-0.11
ST-Linear	0.02	0.16	0.03	0.04
ST-More	0.08	0.06	-0.11	-0.02
ST-Less	0.00	0.00	-0.02	-0.13
ST-Non	0.17	0.12	-0.09	-0.11

This table presents the out-of-sample CERs (in percentage) for the predictive models based on FB and ST. The portfolio weight is restricted in between -1 and 2. Risk aversion is 3 or 10. 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 January 1986 to November 2015.

Appendix C

Figures

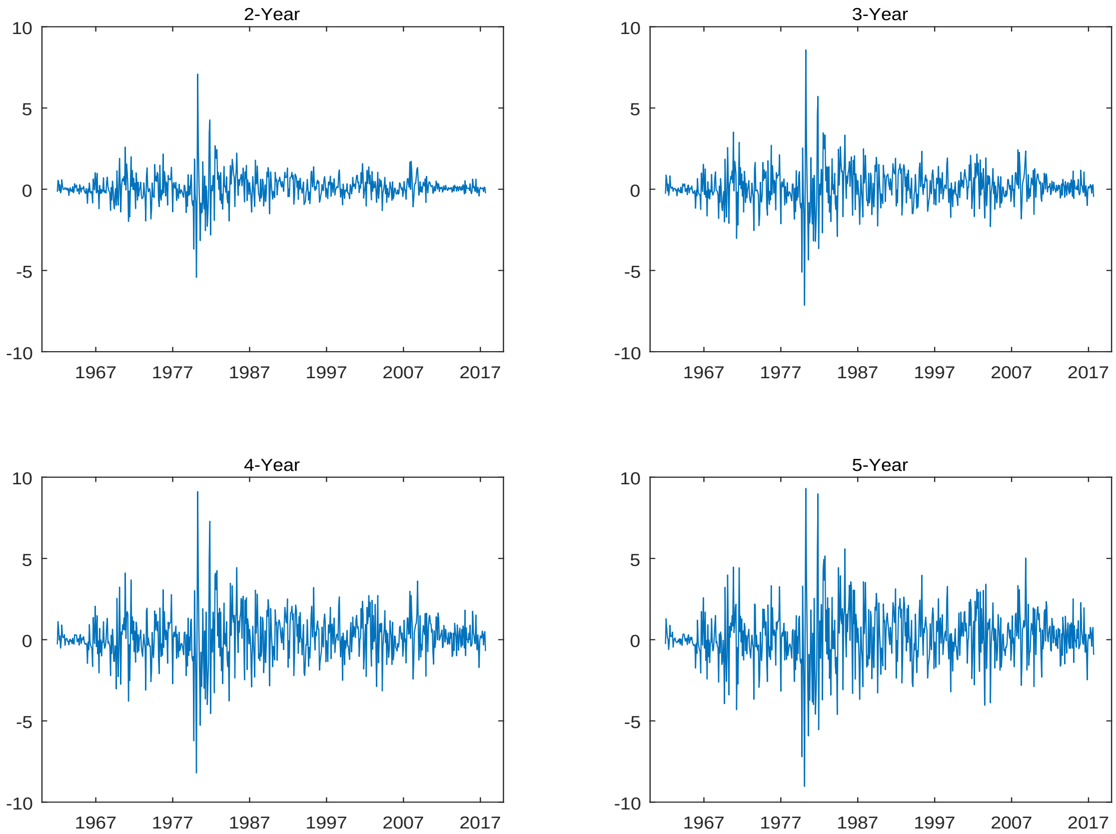

Figure C.1: **The Time Series of Excess Bond Returns**
This figure plots the time series of 4 excess bond returns (in percentage), from Jan, 1962 to Sep, 2017.

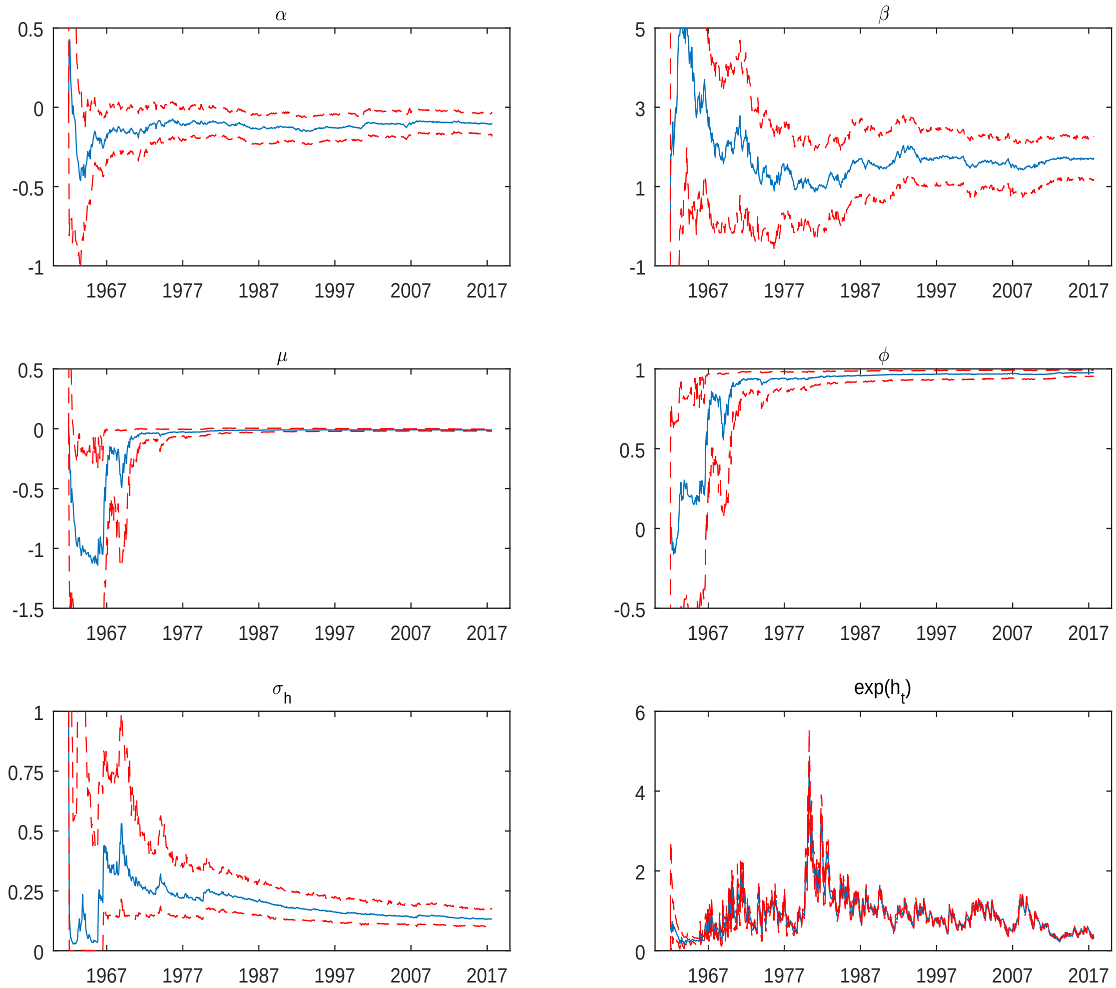

Figure C.2: **Parameter Learning for FB-SV**

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 (2.5) and (2.6). 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.

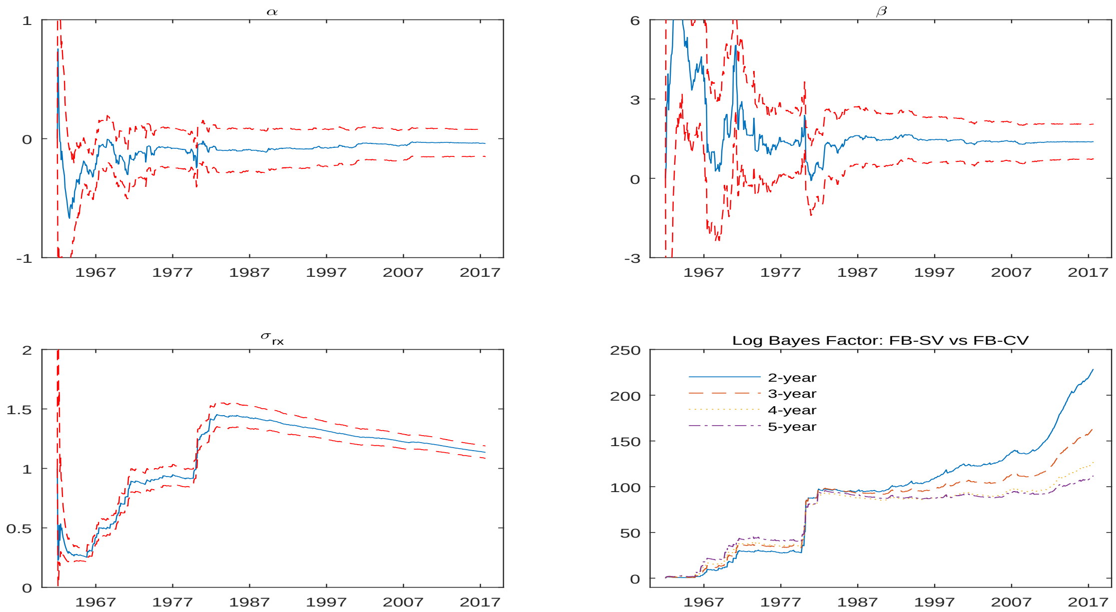

Figure C.3: **Parameter Learning for FB-CV and Bayes Factor**

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 (3.2). 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.

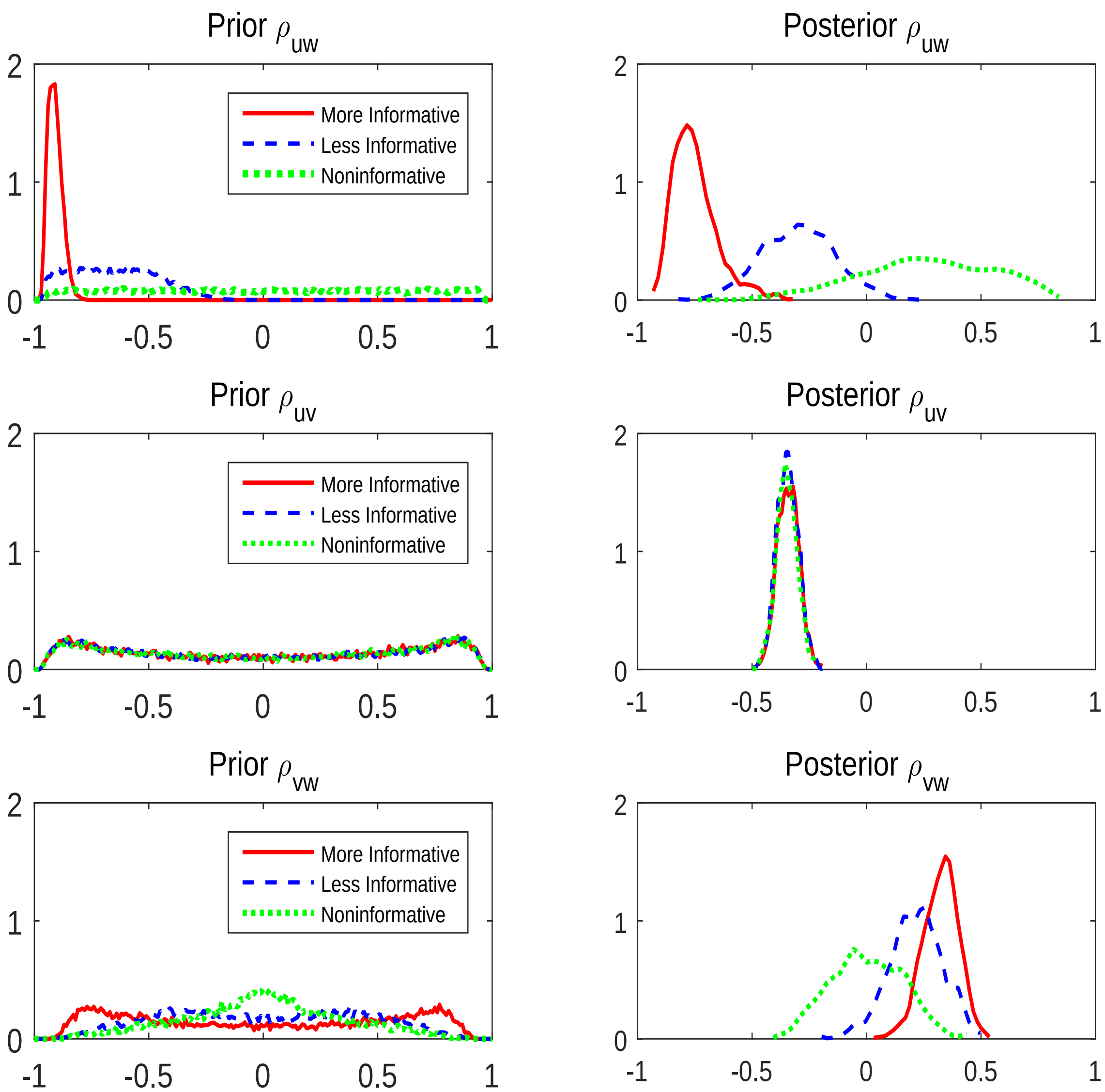

Figure C.4: **The Prior and Posterior Distributions for Correlation Coefficients**
This figure plots the prior and posterior distributions for ρ_{uw}, ρ_{uv}, and ρ_{vw}. The solid, dashed, and dotted distributions represent the distributions associated with the more informative, the less informative, and the noninformative priors for ρ_{uw}.

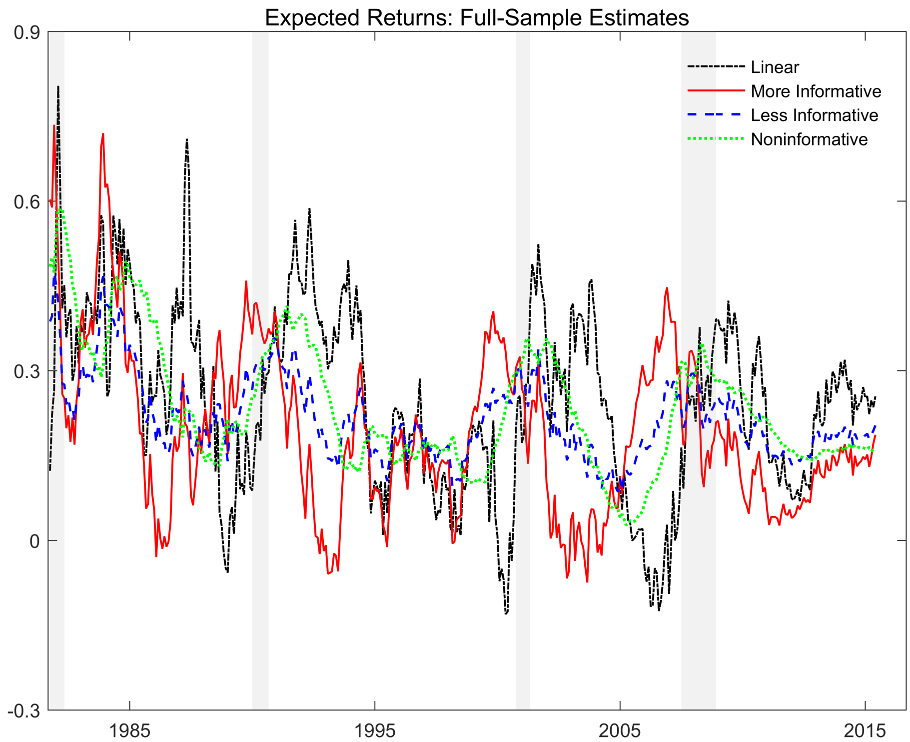

Figure C.5: **Expected Returns Estimates**

This figure plots the full-sample estimates of expected returns (in percentage). The solid, dashed, and dotted distributions represent results from the more informative, the less informative, and the noninformative priors on ρ_{uw}. In this example, the FB predictor is used to forecast 3-year excess return. The shaded area represents NBER recession period. Full sample period is from April 1982 to December 2015.

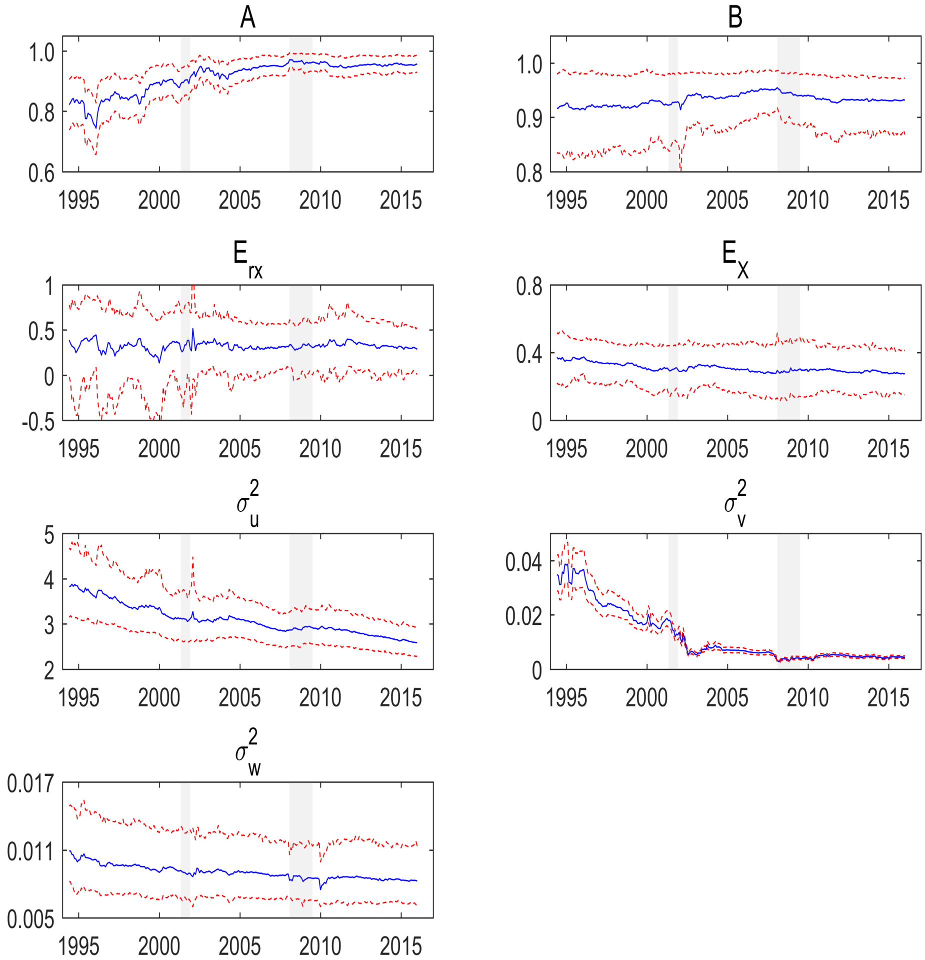

Figure C.6: **Out-of-Sample Parameter Learning Example: Predictive System** This figure plots the parameter learning results from the predictive system with more informative priors. In this example, the CP predictor is used to forecast 5-year excess return. The solid line represents the mean of the distribution. The dashed lines represent the 5th and 95th percentiles. The shaded area represents NBER recession period. Out-of-sample period is from April 1994 to December 2015.

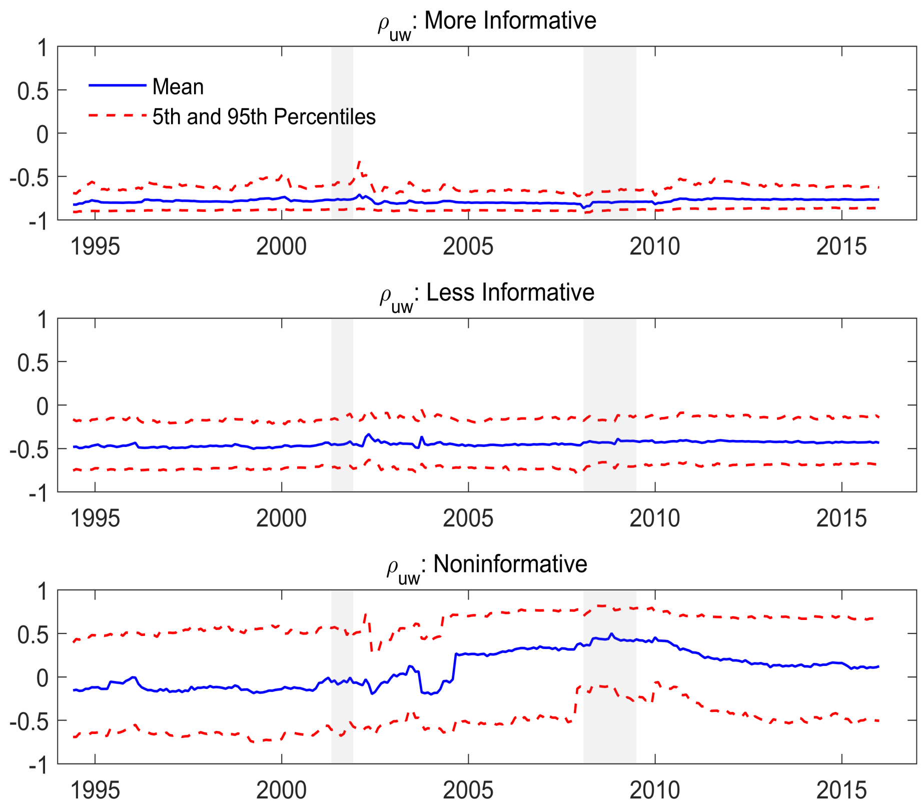

Figure C.7: **Out-of-Sample Parameter Learning Example: Predictive System** This figure plots the parameter learning results for ρ_{uw} (the correlation between expected and unexpected returns) from the predictive system with 3 different priors. In this example, the CP predictor is used to forecast 5-year excess return. The solid line represents the mean of the distribution. The dashed lines represent the 5th and 95th percentiles. The shaded area represents NBER recession period. Out-of-sample period is from April 1994 to December 2015.

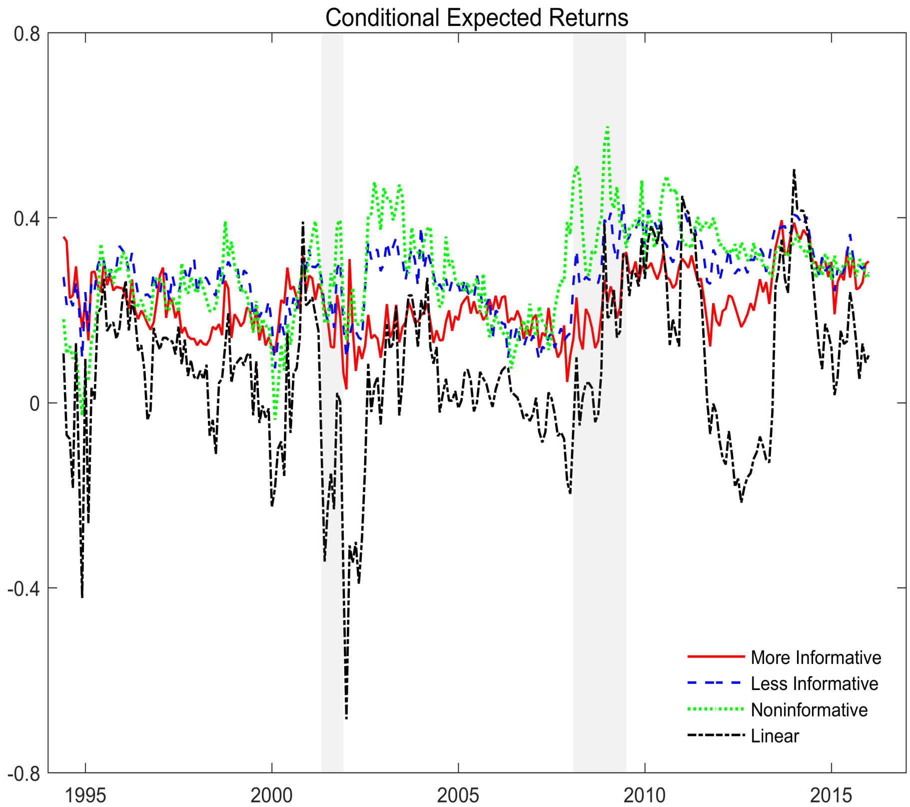

Figure C.8: **Out-of-Sample Conditional Expected Returns**
This figure plots the conditional expected return forecasts (in percentage) for the predictive models in the out-of-sample period. The CP predictor is used to forecast 5-year excess return. The shaded area represents NBER recession period. Out-of-sample period is from April 1994 to December 2015.

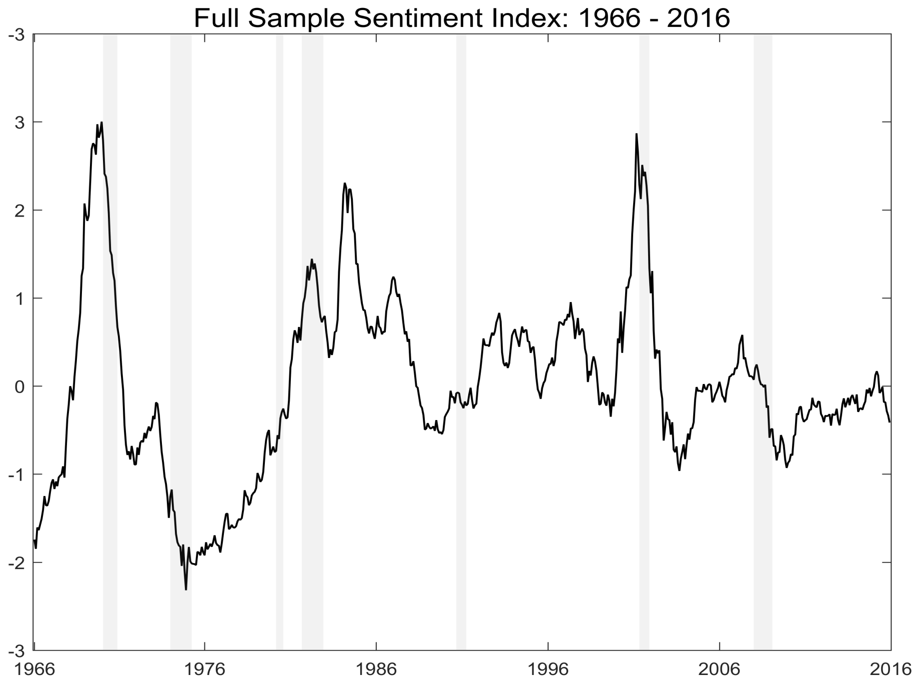

Figure C.9: **Full-Sample Sentiment Index: 1966 - 2016**
The figure shows full sample sentiment index from 1966 to 2016. The shaded area represents NBER recession period.

Appendix D

Bibliography

Bibliography

[1] Baele, L., G. Bekaert, and K. Inghelbrecht. 2010. "The Determinants of Stock and Bond Comovements". *Review of Financial Studies* 23, 374428.

[2] Bai, J., and S. Ng., 2002. "Determining the Number of Factors in Approximate Factor Model". *Econometrica* 70, 135–172.

[3] Baker, M., Wurgler, J., 2006. "Investor sentiment and the cross-section of stock returns". *Journal of Finance* 61, 2219–2257.

[4] Baker, M., Wurgler, J., 2007. "Investor sentiment in the stock market". *Journal of economic perspectives* 21, 129–152.

[5] Baker, M., Wurgler, J., 2012. "Comovement and predictability relationships between bonds and the cross-section of stocks". *Review of asset pricing studies* 21, 57–87.

[6] Bansal, R., and Shaliastovich, I., 2013, "A long-run risks explanation of predictability puzzles in bond and currency markets". *Review of Financial Studies*, 26, 1-33.

[7] Barberis, N., 2000, “Investing for the long run when returns are predictable”. *Journal of Finance* 55, 225–264.

[8] Bauer, M. and Hamilton, J., 2018. “Robust bond risk premia”. *Review of financial studies*, 31, 399–448.

[9] Bekaert, G. and Hodrick, R.J., 2001. “Expectations Hypotheses Tests.” *Journal of Finance* 56, 1357–1394.

[10] 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.

[11] Campbell, J. Y., 1991, “A variance decomposition for stock returns”. *The Economic Journal* 101, 157–179.

[12] Campbell J. Y., Cochrane, J. H. ,1999, “By force of habit: A consumption-based explanation of aggregate stock market behavior”. *Journal of Political Economics* 107, 205–251.

[13] Campbell, J. Y., Shiller, R. J., 1991, “Yield spreads and interest rate movements: A bird’s eye view”, *Review of Economic Studies* 58, 495–514.

[14] 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.

[15] Carter, C. K. and R. Kohn, 1994, "On Gibbs sampling for state space models", *Biometrika* 81, 541 – 553.

[16] Casella, G., and E. I. George, 1992, "Explaining the Gibbs sampler", *The American Statistician* 46, 167–174.

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

[18] Cieslak, A., and P. Povala. Expected returns in Treasury Bonds. *Review of Financial Studies*, 28, 2859–2901.

[19] Clarida, R., Gali, J., and Gertler, M., 1999. "The science of monetary policy: A new Keynesian perspective", *Journal of Economic Literature* 37, 1661–1707.

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

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

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

[23] Creal, D. and Wu, J. C., 2017. "Bond risk premia in consumption-based models". *Working Paper.*

[24] Dai, Q. and Singleton. K., 2003. "Term structure dynamics in theory and reality". *Review of Financial Studies* 16, 631–678.

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

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

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

[28] Duffee, G., 2002. "Term premia and interest rate forecasts in affine models". *Journal of Finance* 57, 405–443.

[29] Eriksen, J., 2017. "Expected business conditions and bond risk premia". *Journal of Financial and Quantitative Analysis* 52, 1667–1703.

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

[31] Farmer, L. E., Schmidt, L., and Timmermann, A., 2018, "Pockets of predictability", *Working Paper*.

[32] Ferson, W. E., Sarkissian, S. and Simin, T. T., 2003, "Spurious regressions in financial economics?" *Journal of Finance* 58, 13931413.

[33] 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.

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

[35] Fulop, A., Li, J., and Wan, R., 2018. Real-time Bayesian learning and bond return predictability. Working Paper.

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

[37] Gargano, A., Pettenuzzo, D., and Timmermann, A.G., 2017. "Bond return predictability: Economic value and links to the macroeconomy". *Management Science*, forthcoming.

[38] Ghysels, E., Horan, C., and Moench, E., 2018. "Forecasting through the Reaview Mirror: Data revisions and bond return predictability". *Review of Financial Studies* 31, 678–714.

[39] Giacoletti, M., Laursen, K., and Singleton, K., 2018. "Learning, dispersion of beliefs, and risk premiums in an arbitrage-free term structure model". *Working Paper.*

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

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

[42] Greenwood, R. and Voyanos, D., 2014. “Bond supply and excess bond returns”. *Review of Financial Studies* 27, 662–713.

[43] 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.

[44] Haddad, V. and Sraer, D., 2018. “The banking view of bond risk premia”. *Working Paper*.

[45] Hansen, L., 2007. “Beliefs, Doubts and Learning: Valuing Macroeconomic Risk”. *American Economic Review* 97, 1–30.

[46] Hanson, S. G., 2014. “Mortgage convexity”. *Journal of Financial Economics*, 113, 270–299.

[47] Hansen, L., 2007. “Beliefs, doubts and learning: Valuing macroeconomic risk”. *American Economic Review* 97, 1–30.

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

[49] Huang, D., Jiang, F., and Tong, G., 2018, “Real time macro factors in bond risk premium”. *Working Paper*.

[50] Huang, D., Jiang, F., Tu, J., and Zhou, G., 2015, “Investor sentiment aligned: a powerful predictor of stock returns”. *Review of Financial Studies* 28, 791–837.

[51] Huang, J. and Shi, Z., 2014. “Determinants of Bond Risk Premia”. *Working Paper*.

[52] Jiang, F. and Tong, G., 2017. “Monetary policy uncertainty and bond risk premium”. *Working Paper*.

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

[54] Johannes, M., Lochstoer, L., and Mou, Y., 2016. “Learning about consumption dynamics”. *Journal of Finance* 71, 551-600.

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

[56] Jurado, K., Ludvgison, S. C., Ng, S., 2015. Measuring uncertainty. *American Economic Review* 105, 1177–1216.

[57] Kelly, B., and S. Pruitt. 2015. "The three-pass regression filter: A new approach to forecasting using many predictors." *Journal of Econometrics* 186, 294–316.

[58] Lin, H., Liu, W., Wu, C., and Zhou, G., 2017. "Prediction of treasury bond returns", working paper.

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

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

[61] Pastor, L., and Stambaugh, R., 2009, "Predictive Systems: Living with Imperfect Predictors". *Journal of Finance* 64, 1583–1628.

[62] Pesaran, M. H., and Timmermann, A., 1995, "Predictability of stock returns: robustness and economic signficance". *Journal of Finance* 50 (4), 1201–1228.

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

[64] 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.

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

[66] Shanken, J. and Tamayo, A., 2012, "Payout yield, risk, and mispricing: A Bayesian analysis". *Journal of Financial Economics* 105, 131–152.

[67] Stambaugh, R., 1999. "Predictive regressions". *Journal of Finance* 54, 375–421.

[68] Taylor, J., 1999. "A historical analysis of monetary policy rules". *Monetary Policy Rules*, University of Chicago Press, Chicago, IL.

[69] 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.

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

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

[72] Wachter, J. A. and Watusawitharana, M., 2015, "What is the chance that the equity premium varies over time? Evidence from regressions on the dividend-price ratio". *Journal of Econometrics* 186, 74–93.

[73] Wan, R., 2018. "Predictive Systems, Real Economy, and Bond Risk Premia". *Working Paper.*

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

[75] Wold, H. 1966. "Estimation of principal components and related models by iterative least squares". In P. R. Krishnaiaah (eds.), Multivariate Analysis, 391420. New York: Academic Press.