University of Glasgow

Cao, Shuo (2016) Essays on the term structure of interest rates. PhD thesis

http://theses.gla.ac.uk/7324/

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Glasgow Theses Service
http://theses.gla.ac.uk/
theses@gla.ac.uk

UNIVERSITY
of
GLASGOW

Essays on the Term Structure of Interest Rates

by

Shuo Cao

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Adam Smith Business School

College of Social Sciences

University of Glasgow

April 2016

Abstract

This PhD thesis contains three main chapters on macro finance, with a focus on the term structure of interest rates and the applications of state-of-the-art Bayesian econometrics. Except for Chapter 1 and Chapter 5, which set out the general introduction and conclusion, each of the chapters can be considered as a standalone piece of work.

In Chapter 2, we model and predict the term structure of US interest rates in a data rich environment. We allow the model dimension and parameters to change over time, accounting for model uncertainty and sudden structural changes. The proposed time-varying parameter Nelson-Siegel Dynamic Model Averaging (DMA) predicts yields better than standard benchmarks. DMA performs better since it incorporates more macro-finance information during recessions. The proposed method allows us to estimate plausible real-time term premia, whose countercyclicality weakened during the financial crisis.

Chapter 3 investigates global term structure dynamics using a Bayesian hierarchical factor model augmented with macroeconomic fundamentals. More than half of the variation in the bond yields of seven advanced economies is due to global co-movement. Our results suggest that global inflation is the most important factor among global macro fundamentals. Non-fundamental factors are essential in driving global co-movements, and are closely related to sentiment and economic uncertainty. Lastly, we analyze asymmetric spillovers in global bond markets connected to diverging monetary policies.

Chapter 4 proposes a no-arbitrage framework of term structure modeling with learning and model uncertainty. The representative agent considers parameter instability, as well as the uncertainty in learning speed and model restrictions. The empirical evidence shows that apart from observational variance, parameter instability is the dominant source of predictive variance when compared with uncertainty in learning speed or model restrictions. When accounting for ambiguity aversion, the out-of-sample predictability of excess returns implied by the learning model can be translated into significant and consistent economic gains over the Expectations Hypothesis benchmark.

Table of Contents

List of Tables

List of Figures

Dedication

To my family.

“*On the road from the City of Skepticism, I had to pass through the Valley of Ambiguity.*”

Adam Smith

Acknowledgements

This PhD thesis is the result of my work at the University of Glasgow over the last three years of study. I am very fortunate to have enjoyed support of many people to whom I am much indebted. Some are particularly worth individual mention with respect to their contribution.

First of all, I am sincerely grateful to my supervisors Professor Dimitris Korobilis and Professor Joseph P. Byrne for their invaluable guidance and insights. Without their generous encouragement and help, I could not strive towards and achieve my goal. Apart from the academic and life advice, I further thank them for offering me precious opportunities and supporting me to pursue an academic career. I also would like to express great gratitude to Dr. Emanuel Moench for his mentorship during my PhD internship at the Deutsche Bundesbank, which is a truly pleasant and fruitful experience. Special thanks go to Professor Gary Koop and Professor Marco Avarucci. It is my pleasure to have them as my thesis examiners, and my viva was a truly pleasant experience.

I acknowledge the full financial support from the College of Social Sciences, University of Glasgow, as well as the conference grant from the Royal Economic Society. The chapters of this thesis and earlier drafts have been presented at several occasions. I have benefited from comments received at Glasgow Economics PhD Workshops, Annual SGPE Methodology Conferences, the 10th BMRC-DEMS Conference, Inquire-Europe Autumn Seminar 2014, the 25th $(EC)^2$ Conference, the 2015 Royal Economic Society (RES) Conference, the 18th Annual Conference of the Swiss Society for Financial Market Research, the Cambridge-INET Economic and Econometric Applications of Big Data Conference, the 6th European Seminar on Bayesian Econometrics, the 13th International Paris Finance Meeting (EUROFIDAI) and the European Winter Meeting of the Econometric Society (EWMES).

I am obliged to Gianni Amisano, Luc Bauwens, Jens Christensen, Giancarlo Corsetti, Richard Crump, Magnus Dahlquist, Pasquale Della Corte, Richard Dennis, Gregory Duffee, Johan Duyvesteyn, Stefano Eusepi, Domenico Giannone, Jan Grobovsek, Kris Jacobs, Eric Leeper, Tatiana Kirsanova, Rajnish Mehra, Theo Nijman, Davide Pettenuzzo, Dooruj Rambaccussing, Barbara Rossi, Christian Schumacher, Bart Taub, Herman van Dijk, Jonathan Wright and Kamil Yilmaz, among many others, for helpful suggestions and discussions. I would further like to thank my colleagues and friends at the Adam Smith Business School, Jilong Chen, Ding Liu, Huichou Huang, Pinho J. Ribeiro, Xiao Zhang, Xuan Zhang, and Yang Zhao with whom I shared not only many interesting and rewarding discussions but also a lot of fun.

A man cannot live without the mental support by his friends and family. My parents and sister Yanyu share all my ups and downs along the way, and their unconditional love and unwavering trust are a permanent source of my power. Finally, a special word of gratitude is reserved for Siting who always stands by me. Everything seems so much more meaningful with her smile.

Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature: ________________

Printed Name: Shuo Cao

CHAPTER 1

Introduction

1.1 General Background

What is the term structure of interest rates? The term structure is the yield curve, which shows interest rates or yields across different maturities (three months, one year, ten years, etc.) at each point in time.[1] For government bonds in a given currency, the term structure can graphically describe the relation between the yield (cost of borrowing) and the time to maturity (see Figure 1.1). Figure 1.1 shows the yield curve, or, in other words, the cross-section of yields, of US Treasury bonds on December 15th, 2015, where the yields are annualized.

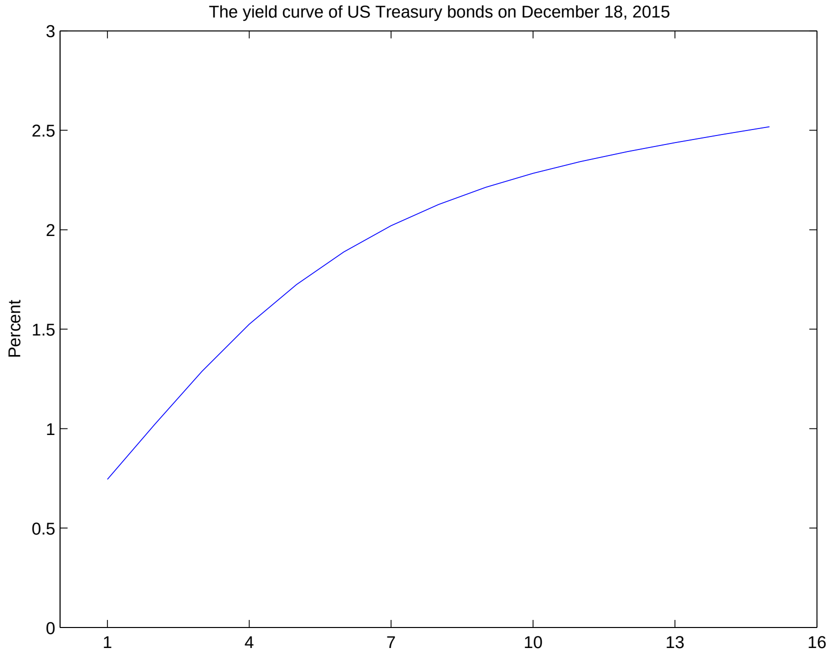
The quest for understanding what moves bond yields has produced a vast amount of literature (see Piazzesi (2010), Gürkaynak and Wright (2012), Duffee (2013) and Diebold and Rudebusch (2013) for extensive reviews). However, current research is still far from perfect, and the literature is quickly expanding. In term structure modeling, *affine term structure models* have gained great popularity. *Affine term structure models*, in a general sense, include models in which bond yields are affine, i.e. linear, in the state vector. Formally, the cross-section of yields can be described by the following equation

$$y(\tau) = A(\tau) + B(\tau)x$$

where the yield of a τ-period bond is denoted as $y(\tau)$, coefficients $A(\tau)$ and $B(\tau)$ depend on maturity τ, and x is the state vector. The cross-sectional restrictions control how bond prices are determined by the state vector, and hence the underlying dynamics are denoted as 'pricing dynamics'. Alternatively, we can denote the dynamics as 'risk-neutral dynamics' in models with no-arbitrage restrictions, as bonds are priced under a so-called 'risk-neutral probability measure'. At their core, *affine term structure models* are factor models, and a low

[1] In finance, maturity refers to the period of time for a financial instrument, at the end of which the financial instrument will cease to exist and the principal is due.

Figure 1.1: *The Term Structure of US Treasury Bonds*



Notes: This figure shows the term structure of US Treasury bonds of maturities 1 year to 15 years on December 15th 2015. The yields are annualized and the unit is percentage.

dimension of states or factors can characterize the vast majority of cross-sectional variation. Piazzesi (2010) indicates that over 97% of the variance is attributable to just three pricing factors; Litterman and Scheinkman (1991) call these factors *Level*, *Slope* and *Curvature*, according to how shocks to these factors affect the yield curve.[2]

The *Level*, *Slope* and *Curvature* factors, as the names imply, govern the level, slope and curvature of the yield curve, respectively. An increase in the *Level* factor increases all yields equally and shifts the curve parallels. An increase in the *Slope* factor increases short yields more than long yields, as bond yields of shorter maturities are loaded more heavily on this factor. For the *Curvature* factor, the medium-maturity yields are loaded most heavily when compared with short yields or long yields, and hence additional variance can be captured.

The predictability of these factors is of great interest, as it is closely related to the expectations of future short rates and risk premia. The time-series properties of pricing factors, which are called the 'historical dynamics' or 'physical dynamics' in term structure modeling, are usually characterized by vector autoregression (VAR) models. To reveal the rich implications of term structure models, we need to consider both the pricing dynamics (cross-section) and the physical dynamics (time series), and a sophisticated term structure model would raise various econometric obstacles in terms of the estimation of these

[2] Alternatively, these factors can be interpreted as factor-mimicking portfolios.

dynamics. This thesis aims to show some promising resolutions to overcome the econometric obstacles by applying Bayesian econometrics.

1.2 Outline of the Thesis

This thesis consists of three main chapters, which are independent but related. This thesis focuses on the term structure of interest rates, i.e., the yield curve, and the applications of Bayesian econometrics. The interest rate term structure is an important topic in macro finance, as it reflects the expectations of market participants about the future path of monetary policy, as well as their assessment of financial market conditions. In this context, this research aims to show that Bayesian econometrics is promising in revealing the true dynamics of yield curves, and provides a new perspective on both macroeconomics and finance.

Chapter 2 proposes a flexible modeling method to incorporate a broad set of conditioning information and to forecast pricing factors that drive the movements of the yield curve. This method, building upon dynamic Nelson-Siegel models, employs a collection of state-of-the-art econometric techniques (dynamic model averaging, time-varying coefficients and stochastic volatility), and significantly outperforms a number of benchmarks in out-of-sample forecasts. In Chapter 3, we study the dynamics of global bond markets by augmenting a Bayesian hierarchical factor model with global macro information. The main finding of this analysis is that the global factor dynamics are stably pinned down by our identification scheme with the macro augmentation. We document the important role of global inflation in driving investors' views on future short rates and risk compensation. We find that latent information of bond yields has economic appeal. Lastly, Chapter 4 explores term structure predictability in an uncertain environment. The Kalman filter and the unscented Kalman filter are utilized to estimate the time-series dynamics and the pricing dynamics, respectively. We find that for a Bayesian agent who can learn from conditional information, the predictive variance mainly comes from the time variation in parameters when compared with the uncertainty in the agent's learning speed and model restrictions. More importantly, an ambiguity-averse investor needs to take into account model uncertainty in order to construct an optimal portfolio that provides significant and consistent economic gains. The following paragraphs provide a brief guide to the main contributions in each of the three chapters. Details on the employed estimation approaches and related empirical results are given in the Appendices.

Chapter 2, we extend the Nelson-Siegel linear factor model by developing a macro-finance framework of the term structure of US interest rates. Our approach is robust to parameter uncertainty and structural change, as we consider instabilities in coefficients and volatilities, and our model averaging method allows for model uncertainty over time. Our time-varying parameter Nelson-Siegel Dynamic Model Averaging (DMA) predicts yields better than standard benchmarks and successfully captures plausible time-varying term premia in real time. We find that the predictability of term structure models tends to be procyclical, while the estimated term premia has a countercyclical pattern. The countercyclicality of term premia is weakened during the financial crisis.

Chapter 2 is structured as follows. The first section includes the introduction and literature review. Section 2.2 describes the framework and the estimation method for modeling bond yield dynamics. Section 2.3 describes the data and discusses the results. Specifically, the first two subsections test the parameter instability and elaborate on the usefulness of employing DMA. Section 2.3.3 displays the point and density forecasting performance of our term structure model. Section 2.3.4 presents the evidence of time-varying predictability and reveals important macro-finance sources that drive the bond yields. Section 2.3.5 shows that the model-implied term premia has informative economic implications. Section 2.4 concludes. Details about the methodology can be found in the Appendices associated with Chapter 2.

Chapter 3 investigates global term structure dynamics using a Bayesian hierarchical factor model augmented with macroeconomic fundamentals. More than half of the variation in the bond yields of seven advanced economies is due to global co-movement. Our results suggest that global inflation is the most important factor among global fundamentals. We evaluate the importance of global inflation by decomposing shocks into a ‘policy channel’ and a ‘risk compensation channel’. Non-fundamental factors are essential in driving global co-movements and are closely related to sentiment and economic uncertainty. Lastly, we analyze asymmetric spillovers in global bond markets connected to diverging monetary policies.

Chapter 3 has the following structure. We discuss the research background and related literature in the beginning of the chapter. In Section 3.2 we introduce the model and describe the estimation and identification of the model. In Section 3.3 we describe the data and present a preliminary data analysis. In Section 3.4 we report empirical results. In particular, we decompose the yield co-movements into two channels and distinguish the role of global inflation. Moreover, we find that non-fundamental factors, which are important in driving global co-movements, are closely related to sentiment and economic uncertainty. Section 3.4.4 sets out the asymmetric ‘spillovers’ in global bond markets. In Section 3.5 we perform robustness checks by testing whether the results are sensitive to the macro spanning condition and zero lower bound. In Section 3.6 we conclude and summarize the implications of this analysis. Technical details about the method, restrictions and identification strategies of factors and shocks appear in Appendix B.2.

Chapter 4 proposes a no-arbitrage framework of term structure modeling with learning and model uncertainty. The representative agent considers parameter instability, as well as the uncertainty in learning speed and model restrictions. We find that apart from observational variance, parameter instability is the dominant source of predictive variance when compared with uncertainty in learning speed or model restrictions. It is important that an ambiguity-averse investor incorporates the ensemble of these salient features to construct the optimal portfolio. We show that accounting for ambiguity aversion, the out-of-sample predictability of excess returns implied by the learning model can be translated into significant and consistent economic gains over the Expectations Hypothesis benchmark.

The structure of Chapter 4 is as follows. The first section discusses the motivation of this chapter and how it builds upon previous literature. Section 4.2 describes the methodology, the term structure models considered, and the framework with ambiguity aversion for evaluating predictability via out-of-sample returns. Section 4.3 reports empirical results on our learning model and the out-of-sample portfolio performance, including a discussion about pricing dynamics, physical dynamics and term structure predictability. Section 4.4 concludes. Technical details are set out in the Appendices. The descriptions of the Kalman filter and the model selection method can be found in Appendix C.2. In particular, I discuss how to implement the unscented Kalman filter in the estimation of the arbitrage-free affine term structure model in Appendix C.4.

CHAPTER 2

Term Structure Dynamics, Macro-Finance Factors and Model Uncertainty

ABSTRACT

This paper models and predicts the term structure of US interest rates in a data rich environment. We allow the model dimension and parameters to change over time, accounting for model uncertainty and sudden structural changes. The proposed time-varying parameter Nelson-Siegel Dynamic Model Averaging (DMA) predicts yields better than standard benchmarks. DMA performs better since it incorporates more macro-finance information during recessions. The proposed method allows us to estimate plausible real-time term premia, whose countercyclicality weakened during the financial crisis.

Keywords: Term Structure of Interest Rates, Nelson-Siegel, Dynamic Model Averaging, Bayesian Methods, Term Premia.

JEL Classification Codes: C32, C52, E43, E47, G17.

Author Contributions: This chapter is drawn from the collaborative work with my supervisors Joseph P. Byrne and Dimitris Korobilis, and a working paper version is available online. I undertook the econometric analysis and did the vast majority of the writing.

2.1 Introduction

Modeling the term structure of interest rates using risk factors is a vast and expanding research frontier in financial economics; see Piazzesi (2010), Gürkaynak and Wright (2012), Duffee (2013) and Diebold and Rudebusch (2013) for extensive reviews. Three pricing factors can capture most of the variation in bond yield data, as indicated in Nelson and Siegel (1987) and Litterman and Scheinkman (1991). Diebold and Li (2006) propose a dynamic Nelson-Siegel (NS) model and successfully predict the yield curve. Our paper builds upon previous work and proposes a term structure model with several novel features. Firstly, to fully capture the factor dynamics, both parameter instability and stochastic volatility in a large system are taken into account. We utilize the dynamic Nelson-Siegel setup with time-varying parameters following Bianchi, Mumtaz and Surico (2009). Our time-varying macro-finance model builds upon a large vector autoregressive (VAR) system with macroeconomic and financial factors in the spirit of Carriero, Kapetanios and Marcellino (2012) and Coroneo, Giannone and Modugno (2015). By extending Koop and Korobilis (2013) a Bayesian method is developed that allows a fast estimation of large systems with many variables.

Secondly, in a reduced-form representation we incorporate financial information in addition to traditional macro variables. Ang and Piazzesi (2003) introduce inflation and the output gap to augment the term structure model and show that macro factors can explain large variation in bond yields. This evidence is echoed by other researchers such as Diebold, Rudebusch and Aruoba (2006), who also stress the importance of key macro variables for the yield curve. Moreover, Moench (2008) shows that a term structure model augmented with a broad macro-finance information set can provide superior forecasts, and the global financial crisis, as an abrupt nonlinear shock, highlighted the importance of the financial market for macroeconomic activity and bond yields more generally. In this paper, we incorporate a substantial range of macro-finance risk factors with modeling techniques that distill large datasets.

Lastly, the proposed model accommodates different degrees of structural changes. Following Koop and Korobilis (2012) we employ Dynamic Model Averaging (DMA) methods in order to determine in a data-based way which macroeconomic or financial risks are relevant for the yield curve. We can choose, at different points in time, between three models: i) one with three pricing factors only; ii) pricing factors plus three key macroeconomic indicators; and iii) pricing factors augmented using up to 15 macro and financial factors. The third macro-finance model is like a 'kitchen sink' model which fully accounts for, and extends, the point of Dewachter and Iania (2012) and Dewachter, Iania and Lyrio (2014) in that financial factors are important for modeling yields, whilst allowing for much more information to be incorporated in the spirit of Ludvigson and Ng (2009). Using DMA probabilities are assigned to each of the models at each point in time and thus averaging is dynamically implemented. When compared with alternative time-varying parameter models, this method is more robust as it encompasses moderate to sudden changes in economic conditions. DMA allows agents to flexibly shift to a more plausible model specification over time, and Elliott and Timmermann (2008) indicate this method can reduce the total forecast risk associated with using only a single 'best' model.

We empirically examine U.S. term structure dynamics using monthly observations from 1971 to 2013. The proposed approach has useful empirical properties in yield forecasting, as it considers parameter and model uncertainty and is robust to potential structural breaks. We compare the forecast performance of DMA to a basic dynamic Nelson-Siegel model and several variants, and show that gains in predictability are due to the ensemble of salient features – time-varying coefficients, stochastic volatility and dynamic model averaging. We find that the predictability of term structure models is time-varying and tends to be procyclical, and macro-finance information is important during recessions. The superior out-of-sample forecasting performance of DMA, especially for short rates, reveals plausible expectations of market participants in real time, and the indicators of real activity and the stock market are particularly helpful in explaining the movements.[1] Using only conditional information, DMA provides successful term premium alternatives to full-sample estimates produced by the no-arbitrage term structure models of Kim and Wright (2005), Wright (2011) and Bauer, Rudebusch and Wu (2014). The estimated term premia has a significant countercyclical pattern, but it appears this pattern is weakened in the global financial crisis possibly because of 'flight-to-quality' demand for US bonds.

This paper is structured as follows. Section 2.2 describes the framework and the estimation method for modeling bond yield dynamics. Section 2.3 describes the data and discusses the results. Specifically, the first two subsections test the parameter instability and elaborate on the usefulness of employing DMA. Section 2.3.3 displays the point and density forecasting performance of our term structure model. Section 2.3.4 presents that the evidence of time-varying predictability and reveals important macro-finance sources that drive the bond yields. Section 2.3.5 shows the model-implied term premia has informative economic implications. Section 2.4 concludes.

[1] This is consistent with Kurmann and Otrok (2013) and Bansal, Connolly and Stivers (2014), who relate the changes in the term structure to news shocks on total factor productivity and asset-class risk, respectively.

2.2 Methods

2.2.1 The Cross-Sectional Restrictions

Following Nelson and Siegel (1987) and Diebold and Li (2006) we assume that three factors summarize most of the information in the term structure of interest rates. The Nelson and Siegel (1987) (NS) approach has an appealing structure that is parsimonious, flexible, and allows for an easy interpretation of the estimated factors. Let $y_t(\tau)$ denote yields at maturity τ, then the factor model we use is of the form:[2]

$$y_t(\tau) = L_t^{NS} + \frac{1 - e^{-\tau\lambda^{NS}}}{\tau\lambda^{NS}} S_t^{NS} + \left(\frac{1 - e^{-\tau\lambda^{NS}}}{\tau\lambda^{NS}} - e^{-\tau\lambda^{NS}} \right) C_t^{NS} + \varepsilon_t(\tau), \tag{2.1}$$

where L_t^{NS} is the "Level" factor, S_t^{NS} is the "Slope" factor, C_t^{NS} is the "Curvature" factor and $\varepsilon_t(\tau)$ is the error term. In the formulation above, λ^{NS} is a parameter that controls the shapes of loadings for the NS factors; following Diebold and Li (2006) and Bianchi, Mumtaz and Surico (2009), we set $\lambda^{NS} = 0.0609$. For estimation purposes, we can rewrite the equation (10) in the equivalent compact form,

$$y_t(\tau) = \mathbf{B}(\tau) F_t^{NS} + \varepsilon_t(\tau),$$

where $F_t^{NS} = \left[L_t^{NS}, S_t^{NS}, C_t^{NS} \right]'$ is the vector of three NS factors, $\mathbf{B}(\tau)$ is the loading vector and $\varepsilon_t(\tau)$ is the error term.

The above Nelson-Siegel restrictions on loadings are cross-sectional restrictions. Feunou et al. (2014) show that the NS model is the continuous time limit of their near arbitrage-free class with a unit root in the pricing dynamics. Joslin, Singleton and Zhu (2011) show that no-arbitrage cross-sectional restrictions cannot improve out-of-sample forecasts. In light of their findings, we specify the cross-sectional loadings with NS restrictions and focus on time-series variation of yield factors, in order to improve the forecast performance.[3]

The time series or physical dynamics of factors are augmented with macro-finance information in an unrestricted VAR. In this setup, the macro variables only affect the unobserved NS factors and do not interact contemporaneously with the observed yields, so that they are unspanned by the yields. In other words, a 'knife-edge' restriction is imposed on the coefficients of macro variables in the cross section, while the time-series dynamics are left unconstrained, see Joslin, Priebsch and Singleton (2014) for details.

[2] This is an asymptotically flat approximating function, and Siegel and Nelson (1988) demonstrate that this property is appropriate if forward rates have finite limiting values.

[3] Nevertheless we test the robustness of core results to the no-arbitrage restrictions in Appendix A.3.3.

2.2.2 Yield Factor Dynamics

In the first step, we use a simple ordinary least squares (OLS) to extract three NS factors. We assume these factors are observed without errors, which is a standard assumption in term structure modeling. The interpretation of the Nelson-Siegel factors is of considerable empirical importance. The Level factor L_t^{NS} loads on all maturities evenly. The Slope factor S_t^{NS} approximates the long-short spread, and its movements are captured by placing more weights on shorter maturities. The Curvature factor C_t^{NS} captures changes that have their largest impact on medium-term maturities, and therefore medium-term maturities load more heavily on this factor. In particular, using the setting $\lambda^{NS} = 0.0609$, the C_t^{NS} has the largest impact on the bond at 30-month maturity, see Diebold and Li (2006).[4]

An important and novel aspect of our methodology is in modeling the factor dynamics in the second step. Following Bianchi, Mumtaz and Surico (2009), the extracted Nelson-Siegel factors augmented with macroeconomic variables follow a time-varying parameter vector autoregression (TVP-VAR) of order p of the form

$$\begin{bmatrix} F_t^{NS} \\ M_t \end{bmatrix} = c_t + B_{1t} \begin{bmatrix} F_{t-1}^{NS} \\ M_{t-1} \end{bmatrix} + \cdots + B_{pt} \begin{bmatrix} F_{t-p}^{NS} \\ M_{t-p} \end{bmatrix} + v_t, \tag{2.2}$$

where c_t are time-varying intercepts, $B_{1t}, ..., B_{pt}$ are time-varying autoregressive coefficients, M_t is a vector of macro-finance risk factors, and v_t is the error term. Following Coroneo, Giannone and Modugno (2015) and Joslin, Priebsch and Singleton (2014), we do not impose any restrictions on the above VAR system.

For the purpose of econometric estimation, we work with a more compact form of Eq. (2.2). We can show that the p-lag TVP-VAR can be written as

$$z_t = X_t \beta_t + v_t, \tag{2.3}$$

where $z_t = \left[L_t^{NS}, S_t^{NS}, C_t^{NS}, M_t'\right]'$, M_t is a $q \times 1$ vector of macro-finance factors, $X_t = I_n \otimes \left[z_{t-1}', ..., z_{t-p}'\right]$ for $n = q+3$, $\beta_t = \left[c_t, vec(B_{1t})', \cdots, vec(B_{pt})'\right]'$ is a vector summarizing all VAR coefficients, $v_t \sim N(0, \Sigma_t)$ with Σ_t an $n \times n$ covariance matrix. This regression-type equation is completed by describing the law of motion of the time-varying parameters β_t and Σ_t. For β_t we follow the standard practice in the literature from Bianchi, Mumtaz and Surico (2009) and consider random walk evolution for the VAR coefficients,

$$\beta_{t+1} = \beta_t + \mu_t, \tag{2.4}$$

based upon a prior β_0 discussed below, and $\mu_t \sim N(0, Q_t)$. Following Koop and Korobilis (2013) we set $Q_t = \left(\Lambda^{-1} - 1\right) cov(\beta_{t-1} | \mathcal{D}_{t-1})$ where $\mathcal{D}_{t-1}$ denotes all the available data at time $t-1$ and scalar $\Lambda \in (0, 1]$ is a 'forgetting factor' discounting older observations.

[4] Further discussion of these factors can be found in Appendix A.2.

The covariance matrix Σ_t evolves according to a Wishart matrix discount process (Prado and West (2010)) of the form:

$$\Sigma_t \sim iW(S_t, n_t), \tag{2.5}$$
$$n_t = \delta n_{t-1} + 1, \tag{2.6}$$
$$S_t = \delta S_{t-1} + f\left(v_t^{'} v_t\right), \tag{2.7}$$

where n_t and S_t are the degrees of freedom and scale matrix, respectively, of the inverse Wishart distribution, δ is a 'decay factor' discounting older observations, and $f\left(v_t^{'} v_t\right)$ is a specific function of the squared residuals of our model and explained in the Appendix C.2.1.

Therefore, we have specified a VAR with drifting coefficients and stochastic volatility which allows for model structural instability and regime changes in the joint dynamics of the NS factors and the macroeconomic and financial factors. In Bayesian inference if Markov Chain Monte Carlo is employed, it will be computationally demanding especially in a recursive forecasting context. Here we extend the methodology of Koop and Korobilis (2013) and conduct a fast estimation scheme to provide accurate results while largely speeding up the estimation procedure. We use what is known as a 'forgetting factor' or 'decay factor' to discount the previous information when updating the parameter estimates; detailed information of our empirical methodology can be found in Appendix C.2.1.

2.2.3 Model Selection

2.2.3.1 Uncertainty about Macro-Finance Factors

This paper argues that the possible set of risk factors relevant for characterizing the yield curve can change over time. We are faced, therefore, with multiple potential yield curve models. Hence, we focus on Eq. (2.3) and work with three different model specifications: small, medium, and large. The small-size (NS) model only contains the three yield factors extracted from the Nelson-Siegel model and zero macro variable, therefore $q = 0$ in Eq. (2.3). The middle-size (NS + macro) model includes, in addition to the Nelson-Siegel factors, Federal Fund Rate, CPI and Industrial Production, so $q = 3$. The large (NS + macro-finance) model includes $q = 15$ macroeconomic and financial variables.

Having three models $\mathcal{M}^{(i)} = 1, 2, 3$, in our model space, we use the recursive nature of the Kalman filter to choose among different models at each point in time. That is, for each t we chose the optimal $\mathcal{M}^{(i)}$ which maximizes the probability/weight

$$\pi_t^{(i)} = f\left(\mathcal{M}_{t-1}^{TRUE} = \mathcal{M}^{(i)} | D_{t-1}\right)$$

under the regularity conditions $\sum_{i=1}^{K} \pi_t^i = 1$ and $\pi_t^i \in [0,1]$, and where $\mathcal{M}_{t-1}^{TRUE}$ is the 'true' model at time $t-1$. We estimate these model weights in a recursive manner, in the spirit of the Kalman filtering approach. We follow Koop and Korobilis (2013) and define the updating step

$$\pi_{t|t}^{(i)} \propto \pi_{t|t-1}^{(i)} p^{(i)} \left(z_t | D_{t-1}\right). \tag{2.8}$$

where the quantity $p^{(i)}\left(z_t|D_{t-1}\right)$ is the time t predictive likelihood of model i, using information up to time $t-1$. This quantity is readily available from the Kalman filter and it provides an out-of-sample measure of fit for each model which allows us to construct model probabilities. In this paper we focus on the predictive likelihoods of the three Nelson-Siegel factors when implementing DMA. The time t prior $\pi_{t|t-1}^{(i)}$ is given by

$$\pi_{t|t-1}^{(i)} = \frac{\left(\pi_{t-1|t-1}^{(i)}\right)^{\alpha}}{\sum_{i=1}^{K}\left[\left(\pi_{t-1|t-1}^{(i)}\right)^{\alpha}\right]} \tag{2.9}$$

where $0 < \alpha \leq 1$ is a decay factor which allows discounting exponentially past forecasting performance, see Koop and Korobilis (2013) for more information. When $\alpha \to 0$ we have the case that at each point in time we update our beliefs with a prior of equal weights for each model. When $\alpha = 1$ the predictive likelihood of each observation has the same weight which is basically equivalent to recursively implementing static Bayesian Model Averaging. For all other values between $(0,1)$ Dynamic Model Averaging occurs. In this paper a sufficiently small value is used for α such that the time t prior is flat, and we will show later this can capture the changing economic conditions and increase the predictive performance.

2.2.3.2 Prior Selection

We define a Minnesota prior for our VAR, which provides shrinkage that could prevent overfitting of our larger models. This prior is of the form $\beta_0 \sim N\left(\mathbf{0}, V^{MIN}\right)$ where V^{MIN} is a diagonal matrix with element V_i^{MIN} given by

$$V_i^{MIN} = \begin{cases} \gamma/r^2, & \text{for coefficients on lag } r \text{ where } r = 1, ..., p \\ \underline{\alpha}, & \text{for the intercept} \end{cases}, \tag{2.10}$$

where p is the lag length and $\underline{\alpha} = 1$. The prior covariance matrix controls the degree of shrinkage on the VAR coefficients. To be more specific, the larger the prior parameter γ is, the more flexible the estimated coefficients are and, hence, the lower the intensity of shrinkage towards zero. As the degree of the shrinkage can directly affect the forecasting results, we allow for a wide grid for the reasonable candidate values of γ: $[10^{-10}, 10^{-6}, 0.001, 0.005, 0.01, 0.05, 0.1]$. The best prior γ is selected dynamically according to the forecasting accuracy each value in the grid generates. That is, following Koop and Korobilis (2013) we select γ for each of the three models $M^{(i)} = 1,2,3$ and for each time period. Details of this Dynamic Prior Selection (DPS) procedure can also be found in the Appendix C.2.4.

In this paper we also need to calibrate some other free parameters: the NS factor parameter

λ^{NS} in Eq. (10), the forgetting factor Λ in Eq. (55), and the decay factor δ in Eq. (54).[5] Regarding the forgetting factor and the decay factor, we follow recommendations in Koop and Korobilis (2013). Intuitively, these parameters control the discounting of past information, which occurs at an exponential rate. When these parameters are equal to one, the model becomes a constant parameter model. Values smaller than one discount past data at a faster rate, allowing faster switches of model parameters. However, too small values may induce sudden changes to outliers, so the state space system is not stable and the results will not be reliable. Hence, following Koop and Korobilis (2013), we choose relatively high values (but less than one) to ensure stability while still allowing for flexibility: The Λ and δ are set to 0.99 and 0.95, respectively.

[5] Following Diebold and Li (2006), Bianchi, Mumtaz and Surico (2009) and Van Dijk et al. (2014) we set $\lambda^{NS} = 0.0609$.

2.3 Data and Results

This study uses the smoothed yields provided from the US Federal Reserve by Gürkaynak, Sack and Wright (2007). We also include 3- and 6-month Treasury Bills (Secondary Market Rate). The empirical analysis focuses on yields with maturities of 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108 and 120 months. The key macroeconomic and financial variables that enter our Dynamic Model Averaging model are obtained from St. Louis Federal Reserve Economic Data (FRED). These include inflation, real activity indicators, monetary policy tools, as well as the stock market, exchange rate, house prices and other financial market indicators; the details can be found in . The full sample is from November 1971 to November 2013 and we use end of the month yield data. The 1, 3, 6 and 12 months ahead predictions are produced with a training sample of 38 observations from the start of our sample, up to and including December 1974. We present the yields' descriptive statistics in Table 3.1. As expected the mean of yields increase with maturity, consistent with the existence of a risk premium for long maturities. Yields have high autocorrelation which declines with lag length and increases with maturity. The short end of the yield curve is more volatile than the long end.

Different numbers of macro-finance variables are selected for the three VARs entering our DMA. As mentioned above, the small-size VAR (NS) does not include any macro or financial variables, but only the Nelson-Siegel factors. The middle-size VAR (i.e. NS + macro) includes Federal Fund Rate, inflation and Industrial Production, which are also used in related literature such as Ang and Piazzesi (2003) and Diebold, Rudebusch and Aruoba (2006). The large VAR (i.e. NS + macro-finance) includes all 15 macro and financial variables, which should comprehensively include the information the market players are able to acquire.

Table 2.1: *Descriptive Statistics of Bond Yields*

	Mean	Std. Dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
3	5.154	3.341	0.010	16.300	0.987	0.815	0.533
6	5.284	3.320	0.040	15.520	0.988	0.827	0.557
12	5.675	3.440	0.123	16.110	0.987	0.842	0.599
24	5.910	3.355	0.188	15.782	0.988	0.858	0.648
36	6.102	3.259	0.306	15.575	0.989	0.868	0.677
48	6.266	3.161	0.454	15.350	0.990	0.873	0.695
60	6.411	3.067	0.627	15.178	0.990	0.876	0.707
72	6.539	2.980	0.815	15.061	0.990	0.877	0.714
84	6.653	2.902	1.007	14.987	0.990	0.878	0.718
96	6.754	2.833	1.197	14.940	0.990	0.878	0.721
108	6.843	2.772	1.380	14.911	0.990	0.878	0.722
120	6.920	2.720	1.552	14.892	0.990	0.877	0.723
Level	7.437	2.379	2.631	14.347	0.989	0.866	0.700
Slope	-2.277	1.940	-5.824	4.522	0.954	0.492	-0.114
Curvature	-1.424	3.222	-8.948	5.282	0.903	0.634	0.369

Notes: This table presents descriptive statistics for monthly yields at 3- to 120-month maturity, and for the yield curve Level, Slope and Curvature factors extracted from the Nelson-Siegel model. The sample period is 1971:11–2013:11. We use following abbreviations. **Std. Dev.**: Standard Deviation; $\hat{\rho}(\mathbf{k})$: Sample Autocorrelation for Lag k.

2.3.1 Evidence on Parameter Instability

In this section we seek to validate the use of time-varying parameter methods. There is a vast selection of different tests of parameter instability and structural breaks in the literature from both a frequentist and a Bayesian perspective; see for example, Andrews and Ploberger (1994), Hanson (2002) and Rossi (2005). McCulloch (2007) suggests a likelihood-based approach to test parameter instability in a TVP model. The limiting distribution of the test statistics may not be standard and, consequently, its critical values need to be bootstrapped. In the spirit of McCulloch (2007), we construct a likelihood-based test on the small VAR system of the factor dynamics, using the 1983-2013 sample. We bootstrap 5000 samples to recover the test statistics following Feng and McCulloch (1996). Based on our test, the null hypothesis that the coefficients of the VAR are constant over time is rejected at 1% significance level, which means employing the TVP-VAR model is appropriate.

However, all the tests mentioned above are in-sample tests and fail to provide evidence concerning out-of-sample instability. Therefore, instead of explicitly specifying a test of parameter instability we follow a different strategy. First, note that in the case of our model specified in Section 2, the constant parameter Nelson-Siegel model can be obtained as a special case of our proposed time-varying specification, that it is nested.[6] Since our ultimate purpose is to obtain optimal forecasts of the yield curve, "testing" for parameter instability can conveniently boil down to a comparison of predictability between the TVP-VAR and a constant parameter VAR. We employ the test proposed by Diebold and Mariano (1995) and evaluate the predictability of competing models across four forecast horizons ($h = 1, 3, 6, 12$ months) and at all twelve of our maturities. The p-values of the tests are reported in Table 2.2, which correspond to the test of the null hypothesis that the competing TVP-VAR model has equal expected square prediction error relative to the benchmark forecasting model constant parameter VAR (i.e. Diebold and Li (2006)), against the alternative hypothesis that the competing TVP-VAR forecasting model has a lower expected square prediction error than the benchmark forecasting model. Table 2.2 indicates the TVP-VAR consistently outperforms the constant parameter VAR. The test statistic rejects the null for most of the maturities, and especially at longer forecast horizons, so the time-varying parameter model should be preferred as it can provide more robust estimates.

To highlight the importance of the TVP feature, we set out the persistence of the time-varying physical factor dynamics of the small-size VAR in Figure 2.1. This can be examined by considering the behavior of the eigenvalues. We can detect significant changes in all eigenvalues, which reflects indispensable changes in the persistence of pricing factors over time. The first eigenvalue seem relatively stable, but the mild variation in the eigenvalue would translate into sufficiently large changes in long-term expectations. Another observation is the clear rising trend for the third eigenvalue, which implies the third pricing factor is becoming more persistent. Moreover, we find that the second and third eigenvalues have important changes in near recession periods, which is connected to the shifting dynamics of Slope and Curvature factors. This is evidence of sudden structural changes. As macro-

[6] In particular, as Koop and Korobilis (2013) show, by setting the forgetting and decay factors $\Lambda = \delta = 1$, our model is equivalent to the recursive estimation of a model with constant parameters and volatility.

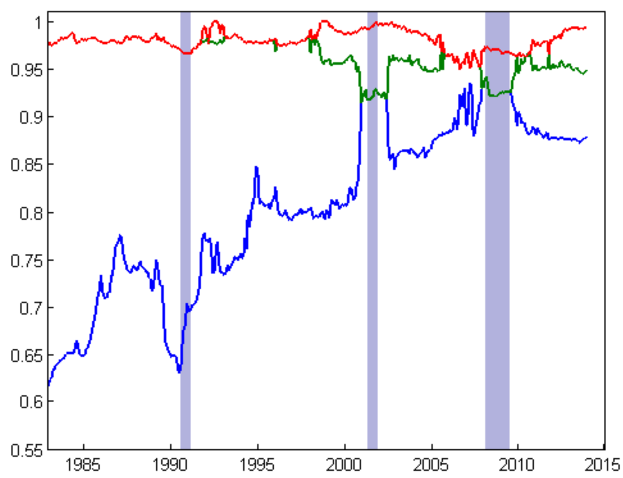
Table 2.2: *Parameter Instability Test*

P-Values: TVP-VAR vs. VAR												
Maturity	3	6	12	24	36	48	60	72	84	96	108	120
$h = 1$	0.02	0.00	0.54	0.14	0.02	0.00	0.00	0.00	0.01	0.08	0.33	0.68
$h = 3$	0.03	0.01	0.13	0.04	0.01	0.01	0.00	0.01	0.02	0.05	0.13	0.28
$h = 6$	0.00	0.00	0.04	0.02	0.01	0.01	0.01	0.01	0.02	0.04	0.08	0.16
$h = 12$	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.03

Notes: 1. This table reports the statistical significance for the relative forecasting performance, based on the Diebold and Mariano (1995) test. We conduct 1, 3, 9 and 12 months ahead forecasts for bond yields at maturities ranging from 3 months to 120 months. The predictive period is between 1983:10 and 2013:11.
2. Statistical significance for the relative MSFE statistics is based on the p-value for the Diebold and Mariano (1995) statistic; the statistic corresponds to the test of the null hypothesis that the competing TVP-VAR model has equal expected square prediction error relative to the benchmark forecasting model constant parameter VAR (i.e. Diebold and Li (2006)), against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the benchmark forecasting model.

finance information is considered important during recessions as suggested by Bernanke, Gertler and Gilchrist (1996), it is uncertain whether the small-size VAR can still produce plausible forecasts when faced with structural instability.

Figure 2.1: *Time-Varying Persistence of Physical Dynamics*



Notes: The graph shows the largest three eigenvalues of the physical dynamics in the small-size TVP model. The shaded areas are recession periods according to the NBER Recession Indicators.

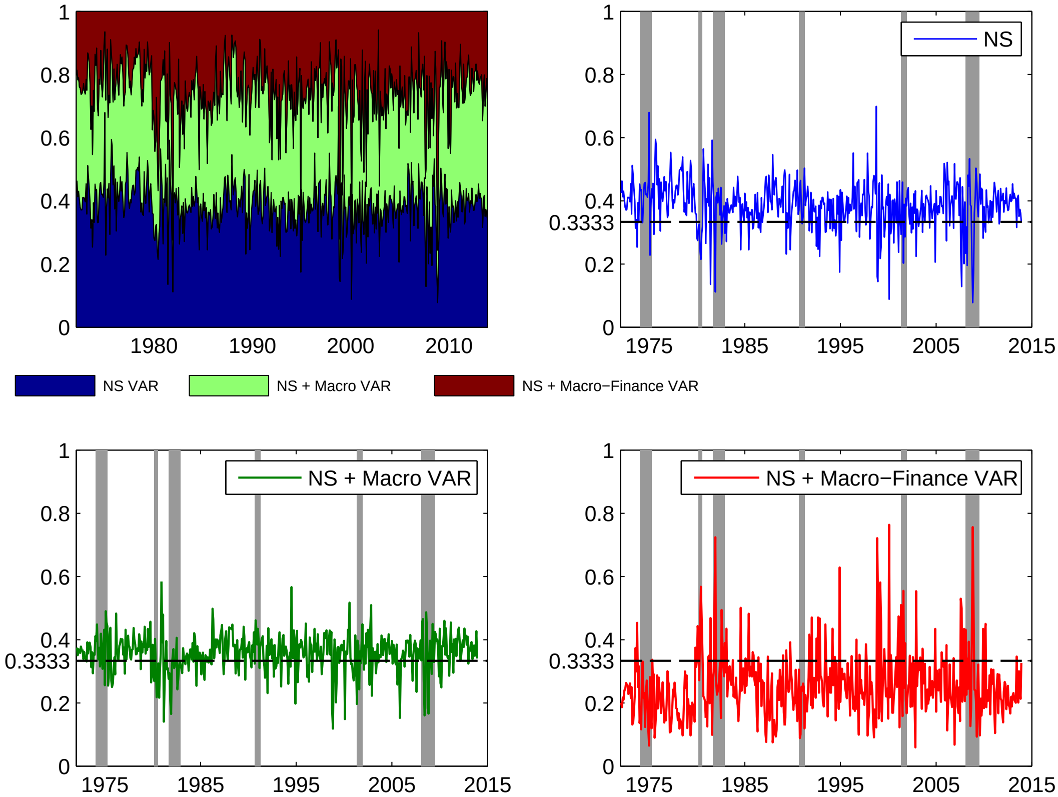
2.3.2 Model Dynamics

In our Bayesian empirical analysis of the factor dynamics, we begin by selecting priors with Dynamic Prior Selection (DPS), then the best prior will be selected for each of the three VAR models. Next we update the model weights with Dynamic Model Averaging (DMA), and finally we update on the parameters using a Bayesian Kalman filter.

In the Dynamic Prior Selection step, we find that the best prior γ value in Eq. (2.10) is stable, i.e. fixed at 0.1, for all three VAR models, given the associated forgetting factor fixed. The associated forgetting factor controls the persistence of probabilities, and the results do not change substantially as long as it is sufficiently large: the best γ values is relatively stable for all three sizes of models when the forgetting factor is larger than 0.90. The evidence concludes that a relatively flexible and consistent prior can generate more accurate yield forecasts. For simplicity and tractability, we fix the value at $\gamma = 0.1$, and therefore the DPS procedure could be skipped in the following analysis. In fact, we find that holding γ constant at 0.1 slightly improves the forecasts, possibly because of the fact that fixing γ reduces posterior parameter uncertainty which in turn can affect uncertainty of posterior predictive densities.

Graphical evidence of the usefulness of our model averaging approach is provided by the Figure 2.2. The upper two panels set out the relative importance of the small, medium and large VAR models used in DMA. In general, there is substantial time variation in the weights, and the empirical observations are of economic importance.

Figure 2.2: *Model Weights for NS, NS plus Macro and NS plus Macro-Finance VAR Models*



Notes:

1. This figure sets out the time-varying probabilities of our three models in our Dynamic Model Averaging (DMA) approach. The probabilities for DMA are updated from a Kalman filter based on the predictive accuracy, see Eq. (70); the probabilities/weights of the VAR models sum up to 1.

2. The upper left panel shows the probability weights of all models. The upper right and the lower panels display the weights of the NS VAR, NS + Macro VAR and the NS + Macro-Finance VAR, respectively. The shaded areas are the recession periods based on NBER Recession Indicators.

Firstly, during recession periods, the approach tends to use more macro-finance information to generate forecasts. The probability of the large-size (macro-finance) model rose steeply and then stayed at a high level during macroeconomic recessions. This is indicated by the higher weights for the macro-finance model during recession periods in the lower right panel of Figure 2.2. In times of acute economic stress, macroeconomic and financial risk factors become more relevant for modeling yields, which is supported by the 'financial accelerator' argument of Bernanke, Gertler and Gilchrist (1996). Among the three, the macro-finance model displays the largest variability in terms of the assigned weights. Hence the additional macro-finance information used to predict yields is appropriately modeled using the DMA approach.

Additionally, the allocated weights of small-size NS model are similar to the medium-size (NS + macro) model. These two models generally have higher weights in the DMA during non-recession periods, but the medium-size model tends to be more stable. This means parsimonious yield curve models with macroeconomic variables, such as Ang and Piazzesi (2003) and Diebold, Rudebusch and Aruoba (2006), are generally effective except during recession periods.

It is worth reiterating the importance of the large macro-finance VAR, as Altavilla, Giacomini and Ragusa (2014) indicate that the original version of the dynamic NS model without macro information has weaker predictive power in recent years. We show the large-size VAR significantly boosts the forecast performance because of its superior performance during the recession periods. Moreover, model averaging expands the model set when compared with a single-model setup or model selection, and potentially mitigates the misspecification problem. Intuitively, the consideration of models with richer information allows us to effectively 'hedge' the risk of using a single model as Elliott and Timmermann (2008) suggest.

Since the changes in model weights are very sensitive to new information, DMA allows us to react to sudden, rather than smooth, changes in coefficients. Without model averaging or selection, a time-varying parameter model with a specific information set may have volatile performance in forecasting, as the true dynamics may not be well captured during certain periods. Our approach encompasses moderate to sudden changes in the economic environment and accordingly is promising in producing more stable forecasting performance.

2.3.3 Forecasting Performance

We now consider the forecasting performance of our approach. We use the Dynamic Model Averaging (DMA) model to predict the yields in a two-step estimation procedure. The first stage is using the Kalman filter to generate predictions of the three Nelson-Siegel yield factors with macro variables, with the addition of DMA. That is, we use Eq. (2.3) with the predicted β_{t+1} to forecast our factors. The second stage is forecasting the yields with the predicted NS factors and the fixed NS loadings. The macro variables are not directly used to predict the yields in the second step, because of the consideration of unspanned macro risks. The predictive duration is from 1983:10 to the 2013:11.

To better evaluate the predictive performance of DMA, we have the following seven variants of dynamic Nelson-Siegel models: recursive estimation of factor dynamics using standard VAR following Diebold and Li (2006) (DL), 10-year rolling-window VAR estimations (DL-R10), recursive VAR estimation with three macro variables (DL-M), recursive estimations of standard VAR with macro-finance principal components following Stock and Watson (2002) (DL-SW), time-varying parameter VAR estimations of factor dynamics without macro information (TVP), time-varying parameter VAR estimations of factor dynamics with three macro variables (TVP-M), and Dynamic Model Selection (DMS).

DL is the two-step forecasting model proposed by Diebold and Li (2006), which recursively estimates the factor dynamics using a standard VAR. In other words, DL estimates the VAR model of factors recursively with historical data, extending through all the following periods. We have four variations of the DL model: 10-year rolling-window estimations (DL-R10); recursive estimations with three macro variables of Fed Fund Rate, Inflation and Industrial Production (DL-M); and recursive estimations with three principal components of our whole macro-finance dataset (DL-SW). In the DL-SW model, three macro principal components are drawn using the method proposed by Stock and Watson (2002) to augment DL. Lastly, we include two extensions of DL using a time-varying parameter VAR without macro information and a time-varying parameter VAR with three macro variables to characterize the factor dynamics, denoted TVP and TVP-M, respectively; the latter is essentially the model estimated in Bianchi, Mumtaz and Surico (2009) using MCMC methods. We report the performance of all models relative to the Random Walk (RW) model so that we can evaluate whether the term structure models successfully capture the high persistence in bond yields.

We assess all models' predictive properties in Table 2.3 which displays the one-period and three-period ahead Mean Squared Forecasting Error (MSFE) Performance for all forecasting models. The core empirical results are very encouraging for the proposed method. As can be seen in Table 2.3, our preferred DMA model consistently outperforms all the benchmark models. Table 2.4 shows the DMA is also preferred at relatively long forecast horizons.[7] The cumulative sum of predictive log-likelihood is displayed in Figure 2.3. It shows that the predictive density of the DMA is more accurate compared to the predictive density of the Diebold-Li (DL) across all maturities, especially for short rates.

Among all models, the results indicate DMA is the only one comparable in forecasting performance to, or better than, the RW. In fact, DMA not only successfully captures the persistence in bond yields, but also reveals robust short rate expectations and risk premium estimates because of its superior performance in short rate forecasts. It is worth noting that the rolling-window forecasts perform much less favorably. In addition, the predictability of DL-SW is not satisfactory. The macro principal components alone cannot provide useful information in terms of yield forecasting, since the method fails to exclude irrelevant information in a time-varying manner. That is, the common information in macro-finance variables may not be useful in forecasting. Hence this result indicates the relative advantages of DMA as a plausible shrinkage method.

In the Nelson-Siegel setup, the long-term yields are almost exclusively driven by the Level factor which is very persistent and has relatively lower volatility, so long-rate forecasts at longer horizons should be quite stable for capable term structure models. For long yields, the forecast performance of a term structure model should be very close to the random walk if the model successfully captures the high persistence as suggested by Duffee (2011*a*). In contrast, if short yields are anchored by policy rates, this implies short-horizon forecasts of short yields are accurate as long as monetary policy is predictable in the short run. However, without further information, forecasts of short yields at longer forecast horizons deteriorate substantially, given that the monetary policy target or market expectations may shift in the long run. In comparing our results to the existing literature, Diebold and Li (2006) shows the DL beats the RW for forecast horizons up to 12 months before 2000. But

[7]The density forecast performance is also reported in Tables 2.3 and 2.4, the log-likelihood of DMA is systematically the highest among all forecasting models.

Table 2.3: *One-Month and Three-Month Ahead Relative MSFE of Term Structure Models*

	DMA	**DMS**	**TVP**	**TVP-M**	**DL**	**DL-R10**	**DL-M**	**DL-SW**
MA	One-Month Ahead Relative MSFE							
3	**0.706**†	0.781	0.747	0.710	0.848	1.085	0.885	1.417
6	**0.818**†	0.927	0.894	0.908	1.068	1.313	1.130	1.668
12	0.971†	1.031	0.983	1.011	0.930	**0.897**	0.979	1.547
24	**1.000**†	1.075	1.044	1.060	1.064	1.105	1.103	1.461
36	**0.977**†	1.039	1.032	1.026	1.123	1.223	1.144	1.237
48	**0.965**†	1.008	1.016	1.002	1.130	1.266	1.143	1.099
60	**0.965**†	0.996	1.011	0.997	1.116	1.273	1.129	1.051
72	**0.971**†	0.998	1.015	1.006	1.096	1.259	1.114	1.055
84	**0.982**†	1.008	1.026	1.024	1.074	1.226	1.098	1.090
96	**0.996**†	1.023	1.040	1.046	1.052	1.173	1.083	1.139
108	**1.009**†	1.038	1.055	1.068	1.031	1.108	1.068	1.183
120	1.020†	1.050	1.065	1.084	**1.015**	1.043	1.053	1.214
Mean	**0.964**†	1.009	1.008	1.010	1.053	1.162	1.083	1.237
MA	Three-Month Ahead Relative MSFE							
3	**0.765**†	0.873	0.864	0.845	1.105	1.514	1.070	1.795
6	**0.863**†	0.976	0.976	0.997	1.305	1.646	1.283	1.907
12	**0.931**†	1.003	0.997	1.019	1.131	1.231	1.119	1.727
24	**0.988**†	1.046	1.062	1.068	1.255	1.390	1.249	1.537
36	**1.002**†	1.044	1.073	1.060	1.295	1.482	1.292	1.358
48	**1.006**†	1.037	1.069	1.049	1.294	1.528	1.293	1.246
60	**1.006**†	1.032	1.063	1.043	1.269	1.539	1.272	1.196
72	**1.005**†	1.030	1.057	1.041	1.233	1.525	1.239	1.189
84	**1.002**†	1.029	1.053	1.044	1.190	1.488	1.201	1.207
96	**0.999**†	1.031	1.050	1.049	1.146	1.431	1.160	1.238
108	**0.996**†	1.033	1.049	1.055	1.102	1.360	1.120	1.272
120	**0.994**†	1.035	1.048	1.061	1.062	1.283	1.083	1.302
Mean	**0.969**†	1.018	1.035	1.032	1.205	1.449	1.205	1.405

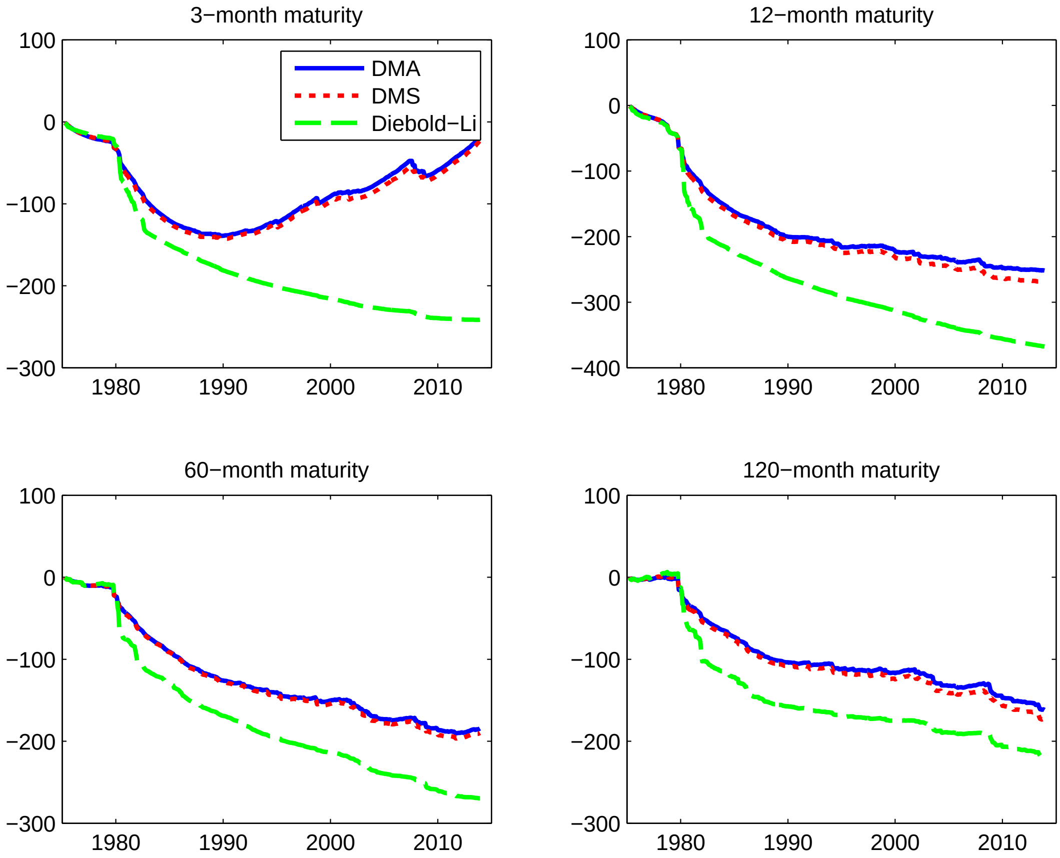
Notes: 1. This table shows 1-month and 3-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from 1983:10 to 2013:11.
2. We report the ratio of each model Mean Squared Forecast Errors (MSFE) relative to Random Walk MSFE, and the preferred values are in bold. The dagger (†) indicates, in terms of the sum of predictive log-likelihood, the model has the preferred value among all models at certain maturities (or in total), see Geweke and Amisano (2010) for details.
3. In this table, we use following abbreviations. **MA**: Maturity (Months); **MSFE**: Mean Squared Forecasting Error; **Mean**: Averaged MSFE across all sample maturities. In our proposed Nelson-Siegel (**NS**) framework, **DMA** (Dynamic Model Averaging) averages all the models with probabilities in each step, while **DMS** (Dynamic Model Selection) chooses the best model with the highest probability at any point in time. **TVP**: a time-varying parameter model without macro information; **TVP-M**: a time-varying parameter model with three macro variables: fund rate, inflation and industrial production, similar to Bianchi Mumtaz and Surico (2009) but estimated with a fast algorithm without the need of MCMC; **DL**: Diebold and Li (2006) model, i.e. constant coefficient Vector Autoregressive model with recursive (expanding) estimations; **DL-R10**: Diebold and Li (2006) estimates based 10-year rolling windows; **DL-M**: factor dynamics in Diebold and Li (2006) are augmented with three macro variables: fund rate, inflation and industrial production, using recursive estimations; **DL-SW**: factor dynamics in Diebold and Li (2006) are augmented with three principal components (see Stock and Watson (2002)) of our macro/finance data, using recursive estimations; **RW**: Random Walk.

Table 2.4: *Relative MSFE Performance of Term Structure Models*

FH	h=6								h=12							
Maturity	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**
3	0.871†	0.976	0.974	1.012	1.332	1.703	1.405	1.908	0.98†	1.073	1.021	1.240	1.349	1.605	1.517	1.677
6	0.947†	1.051	1.053	1.120	1.446	1.796	1.514	1.999	1.034†	1.128	1.079	1.292	1.419	1.703	1.579	1.784
12	0.969†	1.072	1.057	1.080	1.304	1.501	1.322	1.825	1.025†	1.139	1.082	1.210	1.353	1.592	1.458	1.661
24	1.025†	1.109	1.106	1.105	1.393	1.623	1.407	1.707	1.075†	1.191	1.139	1.208	1.474	1.757	1.573	1.664
36	1.038†	1.107	1.110	1.090	1.416	1.685	1.427	1.574	1.091†	1.202	1.152	1.188	1.528	1.848	1.623	1.625
48	1.038†	1.097	1.101	1.073	1.403	1.709	1.414	1.481	1.087†	1.193	1.145	1.163	1.532	1.885	1.625	1.591
60	1.032†	1.085	1.088	1.060	1.368	1.702	1.381	1.432	1.070†	1.175	1.127	1.140	1.505	1.884	1.596	1.577
72	1.021†	1.073	1.076	1.051	1.322	1.673	1.336	1.417	1.049†	1.154	1.106	1.121	1.459	1.858	1.549	1.581
84	1.009†	1.063	1.064	1.046	1.270	1.627	1.286	1.422	1.025†	1.133	1.086	1.105	1.405	1.816	1.494	1.599
96	0.997†	1.055	1.056	1.044	1.218	1.568	1.236	1.438	1.003†	1.115	1.068	1.092	1.349	1.766	1.437	1.623
108	0.987†	1.048	1.049	1.043	1.167	1.502	1.187	1.458	0.983†	1.100	1.054	1.083	1.294	1.711	1.381	1.649
120	0.978†	1.043	1.045	1.044	1.122	1.433	1.142	1.477	0.966†	1.089	1.043	1.077	1.243	1.655	1.329	1.673
Mean	0.994†	1.067	1.067	1.067	1.323	1.632	1.348	1.607	1.035†	1.143	1.093	1.174	1.415	1.748	1.524	1.648

Notes: 1. This table shows six-month and twelve-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from early 1983 to the end of 2013.
2. The MSFE-based statistics relative to the RW are reported. The dagger (†) indicates, in terms of the sum of predictive log-likelihood, the model has the preferred value among all models at certain maturities (or in total), see Geweke and Amisano (2010).
3. In this table, we use following abbreviations. **MSFE**: Mean Squared Forecasting Error; **Mean**: Averaged MAFE across all sample maturities. **DMA** (Dynamic Model Averaging) averages all the models with probabilities in each step, while **DMS** (Dynamic Model Selection) chooses the best model with the highest probability at any point in time. **TVP-M**: a time-varying parameter model with three macro variables: fund rate, inflation and industrial production, similar to Bianchi Mumtaz and Surico (2009) but estimated with a fast algorithm without the need of MCMC; **DL**: Diebold and Li (2006) model, i.e. constant coefficient Vector Autoregressive model with recursive (expanding) estimations; **DL-R10**: Diebold and Li (2006) estimates based 10-year rolling windows; **TVP**: a time-varying parameter model without macro information; **DL-M**: factor dynamics in Diebold and Li (2006) are augmented with three macro variables: fund rate, inflation and industrial production, using recursive estimations; **DL-SW**: factor dynamics in Diebold and Li (2006) are augmented with three principal components (see Stock and Watson (2002)) of our macro-finance data, using recursive estimations; **RW**: Random Walk.

Figure 2.3: *Cumulative Sum of Predictive Log-Likelihood of 3-, 12-, 60- or 120-Month Maturities*



Notes: These are 1-month ahead cumulative sums of predictive log-likelihood for predicted yields from early 1975 to late 2013. From top left clockwise we have maturities of 3, 12, 120 and 60 months. The models are DMA (solid), DMS (dotted) and Diebold-Li (dashed). A higher log-likelihood implies improved density predictability.

Diebold and Rudebusch (2013) and Altavilla, Giacomini and Ragusa (2014) imply NS can no longer beat a RW, which is in line with the increased persistence as we showed previously. Our extended NS model consistently improves upon DL across all horizons and maturities, which is confirmed by Relative MSFEs, predictive log-likelihoods, and the Diebold-Mariano test. Moreover, and at least for shorter horizons, our proposed method improves upon the RW.

Remarks on Predictive Gains Since the pricing dynamics are constrained by the NS restrictions, we can conclude that the predictive gains are purely from the physical dynamics especially by taking parameter and model uncertainty into account. Here we would like to highlight different sources of predictive gains. As mentioned in the last section, the last four columns in Table 2.3 set out the predictive performance of constant-parameter models without stochastic volatility, which are consistently worse than TVP models, no matter whether we include macro information or not. In contrast, our TVP models with stochastic volatility in the third and fourth columns provide significant gains in predictive performance, as they put more weight on the current observations and hence are robust to parameter uncertainty and structural changes.[8] Moreover, introducing an extra layer of model uncertainty is also essential in improving forecast performance. It helps us properly assimilate macro-finance information in a time-varying manner and more importantly, react to abrupt changes, which parallels the 'scapegoat theory' in Bacchetta and Van Wincoop (2004). From the first two columns in Table 2.3, we find further improvement over the TVP models if we allow for both parameter and model uncertainty. Hence, we believe that *the ensemble of these salient features – time-varying coefficients, stochastic volatility and model averaging/selection, is the key to properly incorporate macro-finance information and hence can provide significant gains in predictability*.

To formalize the above arguments, we conduct a statistical test to evaluate the out-of-sample forecasting performance. In Table 2.5 we show results of the Diebold and Mariano (1995) test, in order to evaluate the forecasting performance of DMA relative to DL and TVP-M. The Diebold and Mariano (1995) statistic is also used by Diebold and Li (2006) and Altavilla, Giacomini and Ragusa (2014). The relative MSFE is shown in Table 2.5 for forecasting horizons 1, 3, 6 and 12 months. These results indicate that the DMA clearly outperforms the DL and TVP-M, not only since MSFE are consistently lower but the differences are statistically significant.

2.3.4 Time-Varying Predictability and Macro-Finance Sources

Figure 2.4 shows six-month ahead Squared Forecasting Errors of DL and DMA across the whole out-of-sample forecast period. It is evident that the DMA significantly and consistently outperforms the DL across all maturities. We detect a pattern that the predictability of term structure models, DL in particular, tends to be procyclical. The forecast errors are in general higher during periods when economic conditions deteriorate, especially for short-term rates. Economic theories suggest that central banks can influence short rates to achieve policy goals, so the deteriorated predictability

[8] Additional results about stochastic volatility can be found in Appendix A.3.2.

Table 2.5: *MSFE from DMA Relative to Other Models*

	DMA vs. DL				DMA vs. TVP-M			
Maturity	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
3	0.833***	0.693***	0.653***	0.843***	0.995	0.906*	0.860*	0.790**
6	0.766***	0.661***	0.655***	0.846***	0.901**	0.865**	0.845**	0.800**
12	1.045	0.824**	0.743***	0.866***	0.961**	0.914**	0.897*	0.847**
24	0.939**	0.788***	0.735***	0.849***	0.943***	0.925**	0.927*	0.890*
36	0.870***	0.774***	0.733***	0.845***	0.952***	0.945**	0.952	0.918
48	0.854***	0.777***	0.740***	0.842***	0.963**	0.959*	0.967	0.934
60	0.864***	0.793***	0.754***	0.844***	0.967**	0.965*	0.973	0.939
72	0.886***	0.815***	0.773***	0.846***	0.965**	0.965*	0.971	0.936
84	0.914***	0.842***	0.794***	0.849***	0.959**	0.960*	0.965	0.928
96	0.947**	0.872**	0.819**	0.851***	0.951**	0.953**	0.955	0.918
108	0.978*	0.904**	0.845**	0.854***	0.945***	0.944**	0.946	0.907
120	1.004	0.936	0.872*	0.860***	0.941***	0.937***	0.937	0.897

Notes: 1. This table reports MSFE-based statistics of DMA forecasts of bond yields at maturities ranging from 3 months to 120 months, relative to the forecasts of Diebold and Li (2006) (DL) or TVP-M (similar to Bianchi Mumtaz and Surico (2009)). The predictive period is between 1983:10 and 2013:11.
2. Statistical significance for the relative MSFE statistics is based on the p-value for the Diebold and Mariano (1995) statistic; the statistic corresponds to the test of the null hypothesis that the competing DMA model has equal expected square prediction error relative to the benchmark forecasting model (DL or TVP-M) against the alternative hypothesis that the competing forecasting model has a lower expected square prediction error than the benchmark forecasting model. *, ** and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

implies unexpected or abrupt changes in the behavior of policy makers. For long-term yields, the predictability seems more acyclical, as the movements in long yields are affected not only by short rate expectations but also by the expected risk compensation.

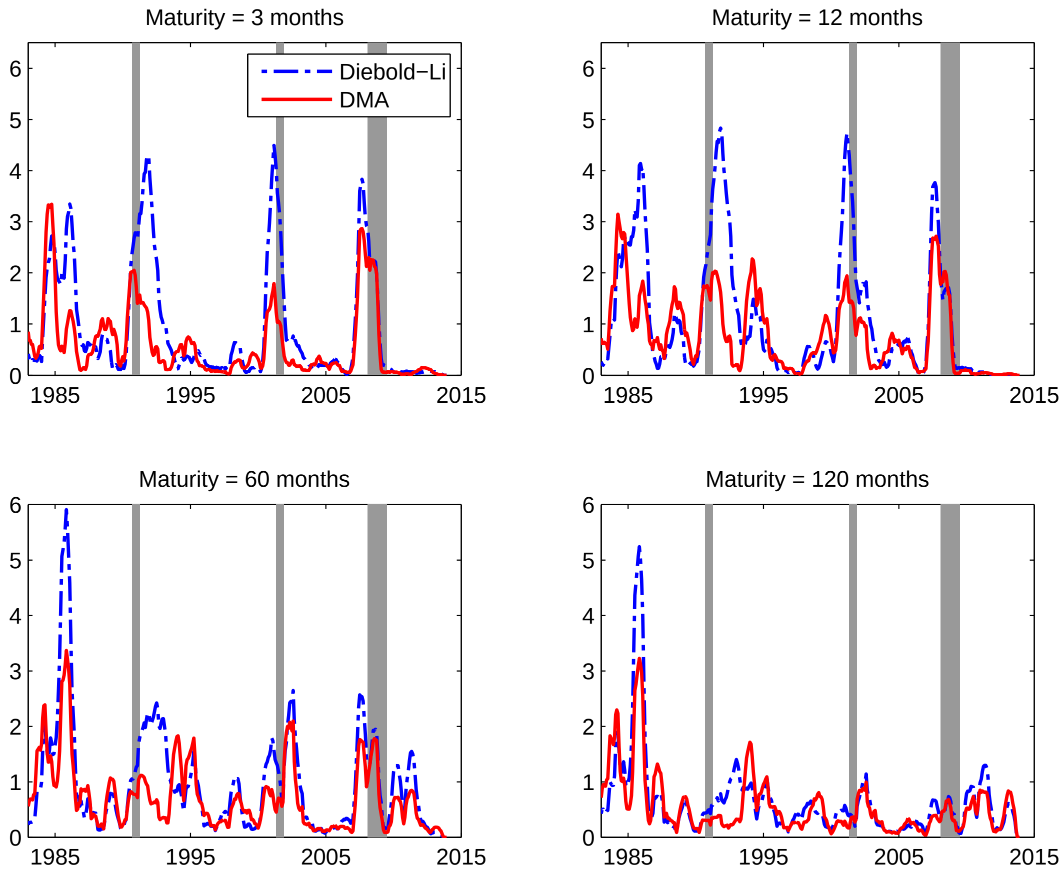
As we have discussed earlier, the DL fails to account for a larger information set and parameter instability, which reduces its forecasting performance. Additionally, our approach allows for model uncertainty, and the large macro-finance VAR significantly contributes to the superior performance of DMA during recession periods. It is of importance to include the large-size VAR, as the increase in the weight assigned to this model significantly reduces forecast errors of DMA when compared with the DL benchmark.[9] Moreover, the DMA has better performance than TVP or TVP-M models especially for short rates as shown in Table 2.3. As we have discussed, DMA allows the model to capture the sudden changes, which in this case are potentially related to the Fed's policy targets.

We are very interested in why the large-size model has distinctive performance during contraction periods. The question is: What are the underlying economic sources that contribute to the pricing factor movements? Following Koop, Pesaran and Potter (1996) and Diebold and Yilmaz (2014), we conduct the *generalized forecast error variance decomposition* to evaluate the contributions of shocks to respective macro-finance variables.[10] Among 15 variables, our results in Figure 2.5 suggest that the most important variables driving large-size VAR predictability are indicators of real activity and the stock market. In particular, real activity and stock markets contribute

[9] The regression results are not shown for the sake of brevity but are available upon request.

[10] We encourage readers to consult the original papers for motivation and background. The generalized variance decomposition is invariant to the ordering of the variables in the VAR, but sums of forecast error variance contributions are not necessarily unity. Here we calculate the normalized weights which add up to unity following Diebold and Yilmaz (2014).

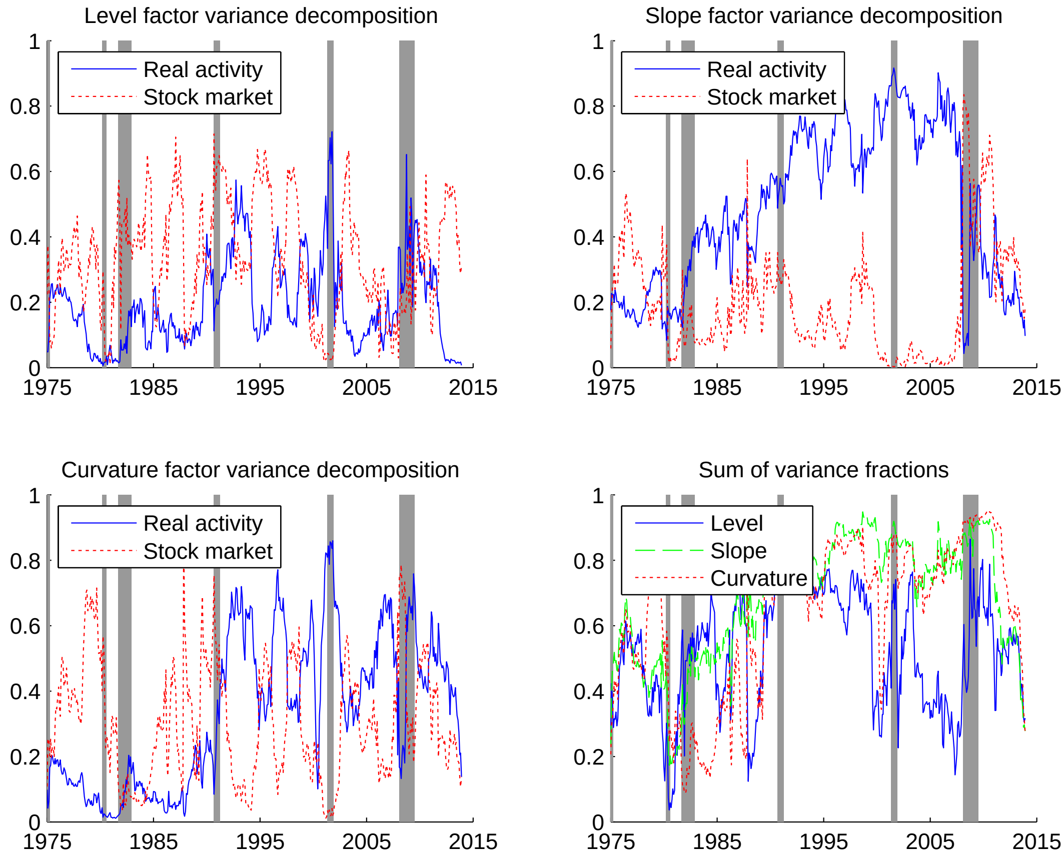
Figure 2.4: *Squared Forecasting Errors for Yields of 3-, 12-, 60- and 120-Month Maturities*



Notes: These are 6 months ahead Squared Forecasting Errors for predicted yields from early 1983 to late 2013. We calculate 9-month moving averages for clarity and plot the statistics for maturities of 3, 12, 60 and 120 months. The models are DMA (solid) and Diebold-Li (dashed and dotted).

to more than 80% of the 60-month forecast error variance of bond factors during the recent three recessions. There is substantial time variation in the role of these variables, and the contributions of two groups tend to be negatively correlated. Specifically, the economic content of Slope and Curvature factors can be largely explained by real activity since the Great Moderation, but the stock market condition is still indispensable. This observation is in line with Kurmann and Otrok (2013) and Bansal, Connolly and Stivers (2014), but contrasts with the evidence from the UK economy provided by Bianchi, Mumtaz and Surico (2009). In the Nelson-Siegel framework, pricing factors are closely related to short rate expectations and term premia, which we will discuss in details in the following.

Figure 2.5: *Variance Decomposition of Bond Pricing Factors*



Notes:
1. This figure sets out the generalized forecast error variance decomposition of pricing factors using the large-size VAR model. The upper panels and the bottom left panel show the average contributions of our target variables to the forecast error variance of the respective bond factors over time. At each point in time, the fractions are calculated based on the 60-month forecast error variance. *Real activity* corresponds to the information of Industrial Production Index and Total Industry Capacity Utilization, and *Stock market* corresponds to the information of S&P 500 Stock Price Index and Wilshire 5000 Total Market Index.
2. The lower right panel displays for each pricing factor the sum of the variance fractions of the two groups of target variables shown in the previous panels. The shaded areas are the recession periods based on NBER Recession Indicators.

Expectation Hypothesis and Term Premium Within our empirical framework we shall set out the formal modeling of the term premia, which has been used to explain the failure of the Expectations Hypothesis and provides important information for the conduct of monetary policy, see Gürkaynak and Wright (2012). The Expectations Hypothesis (EH) consistent bond yield $y_t(\tau)^{EH}$ is given by:[11]

$$y_t(\tau)^{EH} = \frac{1}{\tau}\sum_{i=0}^{\tau-1} E_t y_{t+i}(1), \tag{2.11}$$

where $y_t(\tau)$ is the yield at time t for a bond of τ-period maturity. That is to say, the EH consistent long yield is equal to the average of expected short yields $E_t y_{t+i}(1)$. The time-varying term premium is therefore,

$$TP_t(\tau) = y_t(\tau) - y_t(\tau)^{EH}. \tag{2.12}$$

In the large VAR system, both the short rate expectations and the term premia are linear functions of pricing factors and macro and finance variables, see Diebold, Rudebusch and Aruoba (2006). By the linearity of expectation, we can directly employ the generalized variance decomposition for these quantities.

The patterns in variance decompositions displayed in Figure 2.6 have intuitive appeal, revealing the relative importance of macro-finance variables in driving short rate expectations and risk permia. Standard theory such as the Taylor rule suggests that policy rates should react at least partially to real activity, and our evidence shows short rate expectations are indeed mainly driven by real activity indicators. In contrast, we find that there is strong time variation regarding the main source of risk compensation required by investors, and the underlying sources differ sharply for different horizons. In particular, short-term risk permia is largely explained by real activity shocks during recessions, while long-term risk premia is much less sensitive to real activity during the same periods and more related to the stock market condition in normal times. This observation is interesting but not surprising: As suggested by finance theories, investors' risk attitude influences the demand for bonds and stocks, and Bansal, Connolly and Stivers (2014) show there is a strong link between these two types of assets.

2.3.5 Model-Implied Term Premia

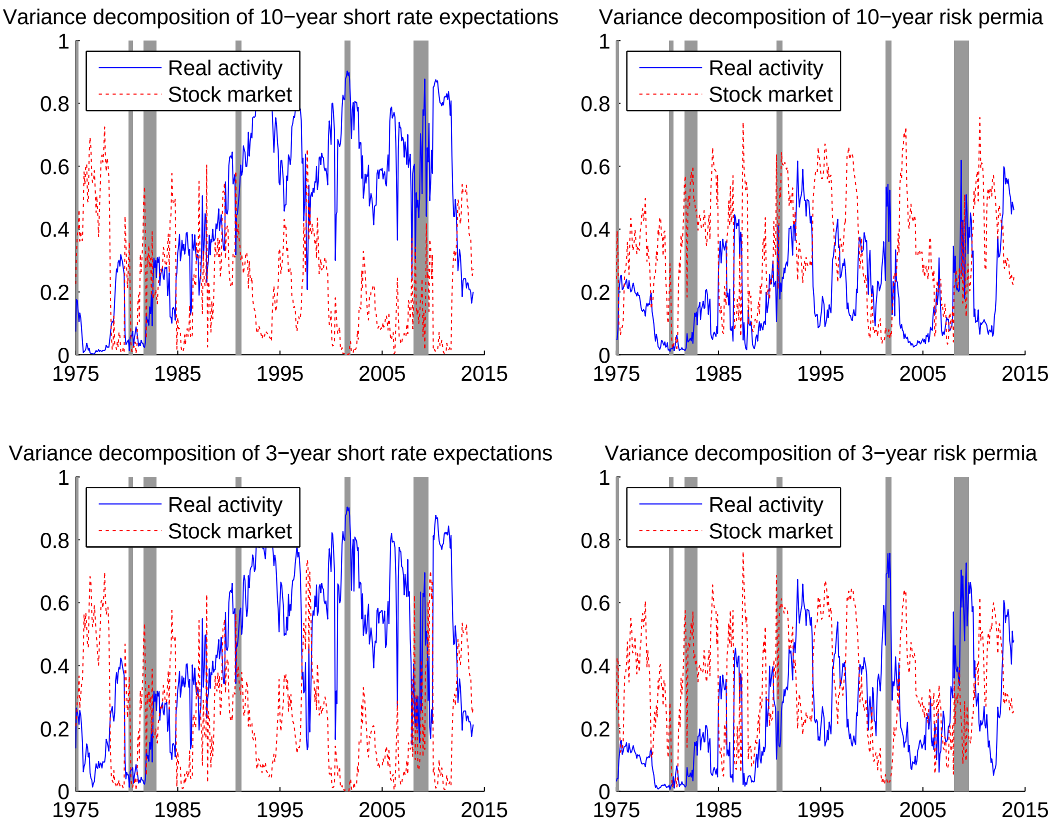
In this section we set out a visual comparison of our term premium estimates. We plot the DMA time-varying risk premia from 1985 for a medium-term bond (maturity 36 months) and a long-term bond (maturity 120 months) in Figure 2.7. For comparison, we also plot the model-implied term premia estimated from no-arbitrage term structure models proposed by Kim and Wright (2005), Wright (2011) and Bauer, Rudebusch and Wu (2014), all of which use full-sample data.[12]

Note that DMA captures plausible term premia using conditional information only. As it is

[11] The expectation here is under the physical measure.

[12] The comparison between the DMA term premia and recursively estimated term premia from dynamic Nelson-Siegel is shown in Appendix A.3.4. The DMA approach seems to be more robust than the constant-parameter dynamic Nelson-Siegel model, as the dynamic Nelson-Siegel model proposed by Diebold and Li (2006) tends to overestimate the future short rates and hence underestimate the term premia.

Figure 2.6: *Variance Decomposition of Short Rate Expectations and Term Premia*

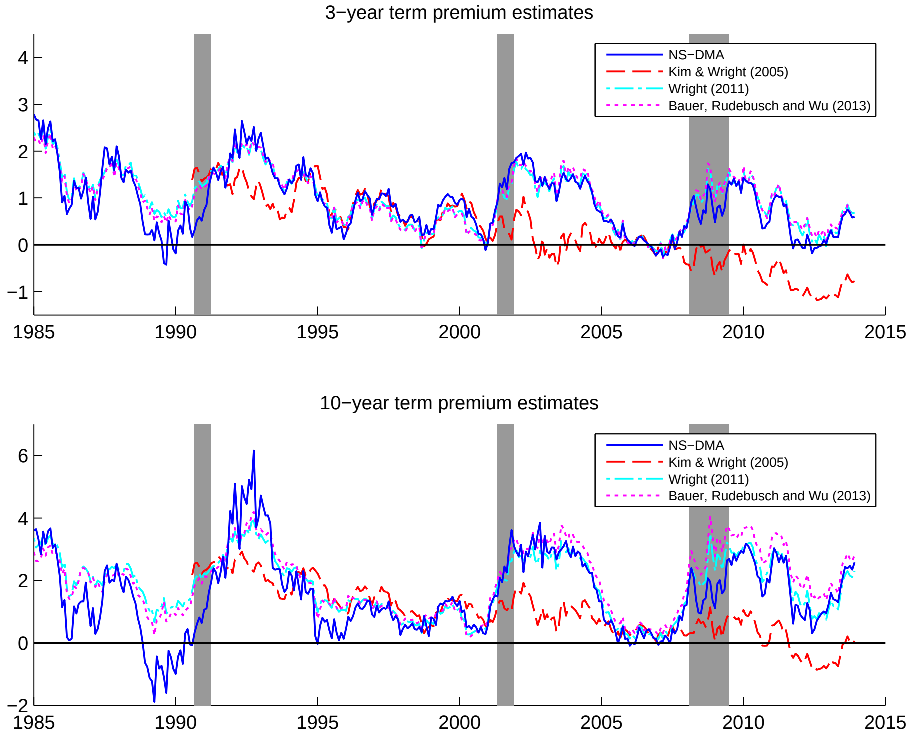


Notes:

This figure sets out the generalized forecast error variance decomposition of short rate expectations and risk premia using the large-size VAR model. The left panels show at each point in time, the average contributions of our target variables to the forecast error variance of 10-year and 3-year short rate expectations, respectively. The right panels show at each point in time, the average contributions of our target variables to the forecast error variance of 10-year and 3-year risk premia, respectively. The time-varying fractions are calculated based on the 60-month forecast error variance. *Real activity* corresponds to the information of Industrial Production Index and Total Industry Capacity Utilization, and *Stock market* corresponds to the information of S&P 500 Stock Price Index and Wilshire 5000 Total Market Index. The shaded areas the recession periods based on NBER Recession Indicators.

shown in the upper panel of Figure 2.7, the 36-month term premium estimates of DMA are highly consistent with the full-sample estimates of Wright (2011) and Bauer, Rudebusch and Wu (2014). In general the term premia displays countercyclical behavior, as they rise in and around US recessions, apart from the estimates of Kim and Wright (2005). The difference between the estimates of Kim and Wright (2005) (KW) and other models is due to the estimated expectation of future short rate. As indicated in Christensen and Rudebusch (2012), there could be potential inaccuracy in the KW measure, because their factor dynamics tend to display much less persistence than the true process. According to the observations here, future short rates from KW would be expected to revert to their mean too quickly, and estimated risk-neutral rates would be too stable, so the KW term premia has a relatively lower variance and may display an acyclical pattern.

Figure 2.7: *Time-Varying Term Premia of 36-and 120-Month Bonds*



Notes:
1. The top panel is the 36-month term premia and the bottom is the 120-month term premia. The EH consistent 36- and 120-month bond yields are estimated using Eq. (4.38); we then calculate the term premia using Eq. (4.41).
2. In addition to DMA, we use the whole sample to separately estimate two types of term premia employing the methods proposed by Wright (2011) and Bauer, Rudebusch and Wu (2014). The Kim and Wright (2005) term premia can be obtained from the Federal Reserve Board website.
3. Shaded areas are recession periods based on the NBER Recession Indicators. The unit is percentage.

Among all measures considered, the DMA term premia seems to be more sensitive to changes in the economic environment, which can be seen more clearly from the lower panel of Figure 2.7 of the long-term term premia. The reason is that expectations of the future short rates move flexibly in DMA and, hence, the 10-year term premia presents a more significant countercyclical pattern. For example, the short rate was continuously decreasing from 1990 to 1993 so the expectation of future short rates was also decreasing. Long rates were relatively stable in contrast, which leads to the increasing risk premia that peaked in 1993.

We can also observe that a divergence between the estimated term premia of DMA and that of Wright (2011) and Bauer, Rudebusch and Wu (2014), lies in the financial crisis period. Christensen, Lopez and Rudebusch (2010) indicate that during the financial crisis, financial markets encountered intense selling pressure because of fears of credit and liquidity risks. The surge in risk aversion creates increased global demand for safe and highly liquid assets, for example, the nominal U.S. Treasury securities. This 'flight-to-quality' could lead to a sharp decline in their yields and therefore result in downward pressure on term premia. Bauer, Rudebusch and Wu (2014) argue, meanwhile, that the procyclical flight-to-quality pressure could not completely offset the usually countercyclical pattern of risk. Based on our estimates, the flight-to-quality demand is evident as shown in the graphs. This makes a distinction between the financial crisis and the previous recessions, as global markets are more unified than ever before and hence capital flows to a safe heaven.

The countercyclical pattern of term premia has been identified in previous literature, such as Estrella and Mishkin (1998), Wright (2006), Kim (2009) and Wheelock and Wohar (2009). D'Agostino, Giannone and Surico (2006) suggest that the term spread may become a weaker indicator of the real economy after the Great Moderation, which parallels the evidence shown in Figure 2.6. In this paper, we present positive evidence that the 'flight-to-quality' demand potentially suppresses the countercyclical pattern of term premia.

2.4 Conclusion

The Nelson-Siegel approach of yield curve modeling has been extended by Diebold and Li (2006), Diebold, Rudebusch and Aruoba (2006) and Bianchi, Mumtaz and Surico (2009). We further extend the literature using a Dynamic Model Averaging (DMA) approach with the consideration of a large set of macro-finance factors, in order to better characterize the nonlinear dynamics of yield factors and further improve yield forecasts. We explore time-varying predictability of term structure models and unfold the time variation of economic sources that drive short rate expectations and risk premia. The DMA method significantly improves the predictive accuracy for bond yields, short rates in particular, and successfully identifies plausible dynamics of term premia in real time. We specifically discuss the countercyclical behavior of term premia and reveal a distinct 'flight-to-quality' demand in the recent financial crisis.

To correctly specify the interactions between the yield factors and macro-finance information, realistic specifications are introduced to enhance this model, such as the settings of unspanned macro risks and time-varying parameters, but these assumptions cause econometric challenges in terms of model tractability. These challenges are addressed here by bringing in a fast and simple estimation technique. The proposed model is believed to be robust, as it is highly consistent with the theoretical and empirical findings in the previous yield curve literature. Future research could employ a one-step approach to provide forecasts with higher accuracy, in which case a trade-off should be made between predictive accuracy and estimation efficiency. Disentangling the real part of the term structure from inflation expectations is meaningful and desirable, but it is beyond the scope of this paper and will be considered for further work.

Bibliography

Abbritti, Mirko, Salvatore Dell'Erba, Antonio Moreno, and Sergio Sola. 2013. "Global factors in the term structure of interest rates." International Monetary Fund International Monetary Fund Working Paper WP/13/223.

Adrian, Tobias, Richard K. Crump, and Emanuel Moench. 2013. "Pricing the term structure with linear regressions." *Journal of Financial Economics*, 110(1): 110–138.

Altavilla, Carlo, Raffaella Giacomini, and Giuseppe Ragusa. 2014. "Anchoring the Yield Curve Using Survey Expectations." European Central Bank Working Paper Series 1632.

Andrews, Donald W.K., and Werner Ploberger. 1994. "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica*, 62(6): 1383–1414.

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

Anh, Le, and Scott Joslin. 2013. "Interest rate volatility and no-arbitrage affine term structure models." University of North Carolina at Chapel Hill Working Paper.

Avramov, Doron. 2002. "Stock return predictability and model uncertainty." *Journal of Financial Economics*, 64(3): 423–458.

Avramov, Doron, and Tarun Chordia. 2006. "Asset pricing models and financial market anomalies." *Review of Financial Studies*, 19(3): 1001–1040.

Bacchetta, Philippe, and Eric Van Wincoop. 2004. "A scapegoat model of exchange-rate fluctuations." *American Economic Review*, 94(2): 114–118.

Bai, Jushan, and Peng Wang. 2015. "Identification and Bayesian estimation of dynamic factor models." *Journal of Business and Economic Statistics*, 33(2): 221–240.

Bai, Jushan, and Serena Ng. 2006. "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions." *Econometrica*, 74(4): 1133–1150.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2013. "Measuring economic policy uncertainty." The University of Chicago Booth School of Business Working Paper.

Bansal, Naresh, Robert A. Connolly, and Chris Stivers. 2014. "The stock-bond return relation, the term-structure's slope, and asset-class risk dynamics." *Journal of Financial and Quantitative Analysis*, FirstView: 1–44.

Bansal, Ravi, and Ivan Shaliastovich. 2010. "Confidence risk and asset prices." *American Economic Review*, 100(2): 537–41.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler. 2005. "Comovement." *Journal of Financial Economics*, 75(2): 283–317.

Bauer, Gregory H., and Antonio Diez de los Rios. 2012. "An international dynamic term structure model with economic restrictions and unspanned risks." Bank of Canada Working Paper.

Bauer, Michael D. 2015. "Restrictions on risk prices in dynamic term structure models." Federal Reserve Bank of San Francisco Working Paper.

Bauer, Michael D., and Glenn D. Rudebusch. 2015. "Resolving the spanning puzzle in macro-finance term structure models." Federal Reserve Bank of San Francisco Working Paper.

Bauer, Michael D., and James D. Hamilton. 2015. "Robust bond risk premia." University of California at San Diego Working Paper.

Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu. 2012. "Correcting estimation bias in dynamic term structure models." *Journal of Business and Economic Statistics*, 30(3): 454–467.

Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu. 2014. "Comment on 'term premia and inflation uncertainty: empirical evidence from an international panel dataset'." *American Economic Review*, 104(1): 323–337.

Benhabib, Jess, and Pengfei Wang. 2015. "Private information and sunspots in sequential asset markets." *Journal of Economic Theory*, 158, Part B: 558 – 584.

Benveniste, Albert, Michel Métivier, and Pierre Priouret. 1990. *Adaptive Algorithms and Stochastic Approximations.* Springer.

Bernanke, Ben, Mark Gertler, and Simon Gilchrist. 1996. "The financial accelerator and the flight to quality." *The Review of Economics and Statistics*, 78(1): 1–15.

Bernanke, Ben S., and Jean Boivin. 2003. "Monetary policy in a data-rich environment." *Journal of Monetary Economics*, 50(3): 525–546.

Bianchi, Francesco, Haroon Mumtaz, and Paolo Surico. 2009. "The great moderation of the term structure of UK interest rates." *Journal of Monetary Economics*, 56(6): 856–871.

Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon. 2012. "Econometric measures of connectedness and systemic risk in the finance and insurance sectors." *Journal of Financial Economics*, 104(3): 535–559.

Björk, Tomas, and Bent Jesper Christensen. 1999. "Interest rate dynamics and consistent forward rate curves." *Mathematical Finance*, 9(4): 323–348.

Bloom, Nicholas. 2014. "Fluctuations in uncertainty." *The Journal of Economic Perspectives*, 28(2): 153–175.

Brandt, Michael W., Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud. 2005. "A simulation approach to dynamic portfolio choice with an application to learning about return predictability." *Review of Financial Studies*, 18(3): 831–873.

Byrne, Joseph P., Giorgio Fazio, and Norbert Fiess. 2012. "Interest rate co-movements, global factors and the long end of the term spread." *Journal of Banking and Finance*, 36(1): 183–192.

Cagetti, Marco, Lars Peter Hansen, Thomas Sargent, and Noah Williams. 2002. "Robustness and pricing with uncertain growth." *Review of Financial Studies*, 15(2): 363–404.

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

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

Carriero, Andrea, and Raffaella Giacomini. 2011. "How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?" *Journal of Econometrics*, 164(1): 21–34.

Carriero, Andrea, George Kapetanios, and Massimiliano Marcellino. 2012. "Forecasting government bond yields with large Bayesian vector autoregressions." *Journal of Banking & Finance*, 36(7): 2026–2047.

Carter, Chris K., and Robert Kohn. 1994. "On Gibbs sampling for state space models." *Biometrika*, 81(3): 541–553.

Chen, Zengjing, and Larry Epstein. 2002. "Ambiguity, risk, and asset returns in continuous time." *Econometrica*, 70(4): 1403–1443.

Christensen, Jens H.E., and Glenn D. Rudebusch. 2012. "The response of interest rates to US and UK quantitative easing." *The Economic Journal*, 122(564): F385–F414.

Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch. 2011. "The affine arbitrage-free class of Nelson-Siegel term structure models." *Journal of Econometrics*, 164(1): 4–20.

Christensen, Jens H.E., Jose A. Lopez, and Glenn D. Rudebusch. 2010. "Inflation expectations and risk premiums in an arbitrage-free model of nominal and real bond yields." *Journal of Money, Credit and Banking*, 42(s1): 143–178.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. "Nominal rigidities and the dynamic effects of a shock to monetary policy." *Journal of Political Economy*, 113(1): 1–45.

Cieslak, Anna, and Pavol Povala. 2014. "Expecting the fed." Northwestern University Working Paper.

Cieslak, Anna, and Pavol Povala. 2015*a*. "Expected returns in treasury bonds." *Review of Financial Studies*, 28(10): 2859–2901.

Cieslak, Anna, and Pavol Povala. 2015*b*. "Information in the term structure of yield curve volatility." *Journal of Finance*, Forthcoming.

Cochrane, John H., and Monika Piazzesi. 2005. "Bond risk premia." *American Economic Review*, 95(1): 138–160.

Cochrane, John H., and Monika Piazzesi. 2008. "Decomposing the yield curve." Stanford University Working Paper.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A. Lochstoer. 2013. "Parameter learning in general equilibrium: The asset pricing implications." National Bureau of Economic Research Working Paper.

Coroneo, Laura, Domenico Giannone, and Michele Modugno. 2015. "Unspanned macroeconomic factors in the yield curve." *Journal of Business and Economic Statistics*.

Coroneo, Laura, Ken Nyholm, and Rositsa Vidova-Koleva. 2011. "How arbitrage-free is the Nelson-Siegel model?" *Journal of Empirical Finance*, 18(3): 393–407.

Cremers, K.J. Martijn. 2002. "Stock return predictability: A Bayesian model selection perspective." *Review of Financial Studies*, 15(4): 1223–1249.

D'Agostino, Antonello, Domenico Giannone, and Paolo Surico. 2006. "(Un)Predictability and macroeconomic stability." European Central Bank Working Paper Series 0605.

Dai, Qiang, and Kenneth J. Singleton. 2000. "Specification analysis of affine term structure models." *The Journal of Finance*, 55(5): 1943–1978.

Dai, Qiang, and Kenneth J. Singleton. 2003. "Term Structure Dynamics in Theory and Reality." *Review of Financial Studies*, 16(3): 631–678.

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

Della Corte, Pasquale, Lucio Sarno, and Daniel L. Thornton. 2008. "The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value." *Journal of Financial Economics*, 89(1): 158–174.

Den Haan, Wouter J., and Steven W. Sumner. 2004. "The comovement between real activity and prices in the G7." *European Economic Review*, 48(6): 1333–1347.

Dewachter, Hans, and Leonardo Iania. 2012. "An extended macro-finance model with financial factors." *Journal of Financial and Quantitative Analysis*, 46(06): 1893–1916.

Dewachter, Hans, and Marco Lyrio. 2008. "Learning, macroeconomic dynamics and the term structure of interest rates." In *Asset Prices and Monetary Policy*. 191–245. University of Chicago Press.

Dewachter, Hans, Leonardo Iania, and Marco Lyrio. 2014. "Information in the yield curve: A Macro-Finance approach." *Journal of Applied Econometrics*, 29(1): 42–64.

Diebold, Francis X., and Canlin Li. 2006. "Forecasting the term structure of government bond yields." *Journal of Econometrics*, 130(2): 337–364.

Diebold, Francis X., and Glenn D. Rudebusch. 2013. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach. The Econometric and Tinbergen Institutes Lectures*, Princeton University Press.

Diebold, Francis X., and Kamil Yilmaz. 2009. "Measuring financial asset return and volatility spillovers, with application to global equity markets." *The Economic Journal*, 119(534): 158–171.

Diebold, Francis X, and Kamil Yilmaz. 2014. "On the network topology of variance decompositions: Measuring the connectedness of financial firms." *Journal of Econometrics*, 182(1): 119–134.

Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13(3): 253–263.

Diebold, Francis X., Canlin Li, and Vivian Z. Yue. 2008. "Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach." *Journal of Econometrics*, 146(2): 351–363.

Diebold, Francis X., Glenn D. Rudebusch, and S. Boroğan Aruoba. 2006. "The macroeconomy and the yield curve: A dynamic latent factor approach." *Journal of Econometrics*, 131(1): 309–338.

Duffee, Gregory R. 2002. "Term premia and interest rate forecasts in affine models." *Journal of Finance*, 57(1): 405–443.

Duffee, Gregory R. 2010. “Sharpe ratios in term structure models.” Department of Economics, Johns Hopkins University Working Paper.

Duffee, Gregory R. 2011*a*. “Forecasting with the term structure: The role of no-arbitrage restrictions.” Johns Hopkins University, Department of Economics Working Paper.

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

Duffee, Gregory R. 2013. “Bond pricing and the macroeconomy.” In *Handbook of the Economics of Finance*. Vol. 2, Part B, , ed. George M. Constantinides, Milton Harris and Rene M. Stulz, 907–967. Elsevier.

Duffee, Gregory R. 2014. “Expected inflation and other determinants of Treasury yields.” Johns Hopkins University, Department of Economics Working Paper.

Duffee, Gregory R, and Richard H Stanton. 2012. “Estimation of dynamic term structure models.” *The Quarterly Journal of Finance*, 02(02): 1250008.

Duffie, Darrell. 2001. *Dynamic Asset Pricing Theory.* Princeton University Press.

Duffie, Darrell, and Rui Kan. 1996. “A yield-factor model of interest rates.” *Mathematical Finance*, 6(4): 379–406.

Eickmeier, Sandra, Leonardo Gambacorta, and Boris Hofmann. 2014. “Understanding global liquidity.” *European Economic Review*, 68: 1–18.

Elliott, Graham, and Allan Timmermann. 2008. “Economic forecasting.” *Journal of Economic Literature*, 46(1): 3–56.

Estrella, Arturo, and Frederic S. Mishkin. 1998. “Predicting US recessions: Financial variables as leading indicators.” *Review of Economics and Statistics*, 80(1): 45–61.

Evans, Charles L., and David A. Marshall. 2007. “Economic determinants of the nominal treasury yield curve.” *Journal of Monetary Economics*, 54(7): 1986–2003.

Evans, George W., and Seppo Honkapohja. 2001. *Learning and Expectations in Macroeconomics.* Princeton University Press.

Fama, Eugene F., and Robert R. Bliss. 1987. “The information in long-maturity forward rates.” *American Economic Review*, 77(4): 680–692.

Feng, Ziding D., and Charles E. McCulloch. 1996. “Using bootstrap likelihood ratios in finite mixture models.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(3): 609–617.

Fernández-Villaverde, Jesús, and Thomas J. Sargent Mark W. Watson Rubio-Ramírez, Juan F. 2007. “ABCs (and Ds) of Understanding VARs.” *The American Economic Review*, 97(3): 1021–1026.

Feunou, Bruno, Jean-Sébastien Fontaine, Anh Le, and Christian Lundblad. 2014. “Term structure modeling when monetary policy is unconventional: A new approach.” Bank of Canada Working Paper.

Filipović, Damir. 1999. “A note on the Nelson–Siegel family.” *Mathematical Finance*, 9(4): 349–359.

Gagliardini, Patrick, Paolo Porchia, and Fabio Trojani. 2009. "Ambiguity aversion and the term structure of interest rates." *Review of Financial Studies*, 22(10): 4157–4188.

Gargano, Antonio, Davide Pettenuzzo, and Allan G. Timmermann. 2014. "Bond return predictability: Economic value and links to the macroeconomy." University of California, San Diego Working Paper.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang. 2007. "Portfolio selection with parameter and model uncertainty: A multi-prior approach." *Review of Financial Studies*, 20(1): 41–81.

Geweke, John, and Gianni Amisano. 2010. "Comparing and evaluating Bayesian predictive distributions of asset returns." *International Journal of Forecasting*, 26(2): 216–230.

Giacoletti, Marco, Kristoffer T. Laursen, and Kenneth J. Singleton. 2014. "Learning, dispersion of beliefs, and risk premiums in an aribitrage-free term structure model." Stanford University Working Paper.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics*, 18(2): 141–153.

Gürkaynak, Refet S., and Jonathan H. Wright. 2012. "Macroeconomics and the term structure." *Journal of Economic Literature*, 50(2): 331–367.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright. 2007. "The US Treasury yield curve: 1961 to the present." *Journal of Monetary Economics*, 54(8): 2291–2304.

Hansen, Lars Peter. 2007. "Beliefs, doubts and learning: Valuing macroeconomic risk." *American Economic Review*, 97(2): 1–30.

Hanson, Bruce E. 2002. "Tests for parameter instability in regressions with I(1) processes." *Journal of Business and Economic Statistics*, 20(1): 45–59.

Hautsch, Nikolaus, and Fuyu Yang. 2012. "Bayesian inference in a stochastic volatility Nelson-Siegel model." *Computational Statistics and Data Analysis*, 56(11): 3774–3792.

Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho. 2011. "What factors drive global stock returns?" *Review of Financial Studies*, 24(8): 2527–2574.

Hubrich, Kirstin, Antonello D'Agostino, et al. 2013. "Financial shocks and the macroeconomy: Heterogeneity and non-linearities." European Central Bank ECB Occasional Paper Series 143.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. "Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software." *PlOS ONE*, 9(6): e98679.

Johannes, Michael, Arthur Korteweg, and Nicholas Polson. 2013. "Sequential learning, predictability, and optimal portfolio returns." *The Journal of Finance*, 69(2): 4.

Joslin, Scott, Anh Le, and Kenneth J. Singleton. 2013. "Gaussian macro-finance term structure models with lags." *Journal of Financial Econometrics*, 11(4): 581–609.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu. 2011. "A new perspective on Gaussian dynamic term structure models." *Review of Financial Studies*, 24(3): 926–970.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton. 2014. "Risk premiums in dynamic term structure models with unspanned macro risks." *The Journal of Finance*, 69(3): 1197–1233.

Jotikasthira, Pab, Anh Le, and Christian T. Lundblad. 2015. "Why do term structures in different currencies comove?" *Journal of Financial Economics*, 115(1): 58–83.

Kim, Chang-Jin, and Charles R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* Vol. 1, the MIT Press.

Kim, Don H. 2009. "Challenges in Macro-Finance modeling." *Federal Reserve Bank of St. Louis Review*, 91(5, Part 2): 519–44.

Kim, Don H., and Jonathan H. Wright. 2005. "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates." Board of Governors of the Federal Reserve System Working Paper.

Kim, Don H., and Kenneth J. Singleton. 2012. "Term structure models and the zero bound: An empirical investigation of Japanese yields." *Journal of Econometrics*, 170(1): 32–49.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit.* Hougton Mifflin Company.

Koop, Gary, and Dimitris Korobilis. 2009. "Bayesian multivariate time series methods for empirical macroeconomics." *Foundations and Trends® in Econometrics*, 3(4): 267–358.

Koop, Gary, and Dimitris Korobilis. 2012. "Forecasting inflation using Dynamic Model Averaging." *International Economic Review*, 53(3): 867–886.

Koop, Gary, and Dimitris Korobilis. 2013. "Large time-varying parameter VARs." *Journal of Econometrics*, 177(2): 185–198.

Koop, Gary, and Dimitris Korobilis. 2014. "A new index of financial conditions." *European Economic Review*, 71(0): 101 – 116.

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter. 1996. "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, 74(1): 119–147.

Koopman, Siem Jan, Max I.P. Mallee, and Michel Van der Wel. 2010. "Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters." *Journal of Business and Economic Statistics*, 28(3): 329–343.

Kose, M. Ayhan, Christopher Otrok, and Charles H. Whiteman. 2003. "International business cycles: World, region, and country-specific factors." *American Economic Review*, 93(4): 1216–1239.

Kozicki, Sharon, and Peter A. Tinsley. 2001. "Shifting endpoints in the term structure of interest rates." *Journal of Monetary Economics*, 47(3): 613–652.

Krippner, Leo. 2015. "A theoretical foundation for the Nelson-Siegel class of yield curve models." *Journal of Applied Econometrics*, 30(1): 97–118.

Kumar, Alok, and Charles Lee. 2006. "Retail investor sentiment and return comovements." *The Journal of Finance*, 61(5): 2451–2486.

Kurmann, André, and Christopher Otrok. 2013. "News shocks and the slope of the term structure of interest rates." *American Economic Review*, 103(6): 2612–32.

Kurz, Mordecai. 1994. "On rational belief equilibria." *Economic Theory*, 4(6): 859–876.

Laubach, Thomas, Robert J. Tetlow, and John C. Williams. 2007. "Learning and the role of macroeconomic factors in the term structure of interest rates." Board of Governors of the Federal Reserve System Working Paper.

Lee, Bong-Soo. 1998. "Permanent, temporary, and non-fundamental components of stock prices." *Journal of Financial and Quantitative Analysis*, 33(01): 1–32.

Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang. 2013. "Fiscal foresight and information flows." *Econometrica*, 81(3): 1115–1145.

Litterman, Robert B., and Jose Scheinkman. 1991. "Common factors affecting bond returns." *The Journal of Fixed Income*, 1(1): 54–61.

Litterman, Robert B., José Scheinkman, and Laurence Weiss. 1991. "Volatility and the yield curve." *The Journal of Fixed Income*, 1(1): 49–53.

Liu, Jun, Jun Pan, and Tan Wang. 2005. "An equilibrium model of rare-event premia and its implication for option smirks." *Review of Financial Studies*, 18(1): 131–164.

Ludvigson, Sydney C. 2004. "Consumer confidence and consumer spending." *The Journal of Economic Perspectives*, 18(2): 29–50.

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

Markowitz, Harry. 1952. "Portfolio Selection." *The Journal of Finance*, 7(1): 77–91.

McCulloch, J. Huston. 2007. "The Kalman foundations of adaptive least squares, with application to US inflation." Ohio State University Working Paper.

Moench, Emanuel. 2008. "Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach." *Journal of Econometrics*, 146(1): 26–43.

Moench, Emanuel. 2012. "Term structure surprises: The predictive content of curvature, level, and slope." *Journal of Applied Econometrics*, 27(4): 574–602.

Moench, Emanuel, Serena Ng, and Simon Potter. 2013. "Dynamic hierarchical factor models." *Review of Economics and Statistics*, 95(5): 1811–1817.

Neftci, Salih. 2004. *Principles of Financial Engineering.* Academic Press Inc.

Nelson, Charles R., and Andrew F. Siegel. 1987. "Parsimonious modeling of yield curves." *Journal of Business*, 60(4): 473–489.

Novy-Marx, Robert. 2014. "Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars." *Journal of Financial Economics*, 112(2): 137–146.

Orphanides, Athanasios, and Min Wei. 2012. "Evolving macroeconomic perceptions and the term structure of interest rates." *Journal of Economic Dynamics and Control*, 36(2): 239–254.

Pástor, L'uboš, and Robert F Stambaugh. 1999. "Costs of equity capital and model mispricing." *The Journal of Finance*, 54(1): 67–121.

Pástor, L'uboš, and Robert F Stambaugh. 2000. "Comparing asset pricing models: an investment perspective." *Journal of Financial Economics*, 56(3): 335–381.

Pettenuzzo, Davide, Allan Timmermann, and Rossen Valkanov. 2014. "Forecasting stock returns under economic constraints." *Journal of Financial Economics*, 114(3): 517–553.

Piazzesi, Monika. 2010. "Affine term structure models." *Handbook of Financial Econometrics*, 1: 691–766.

Piazzesi, Monika, and Martin Schneider. 2007. "Equilibrium yield curves." In *NBER Macroeconomics Annual 2006, Volume 21*. 389–472. MIT Press.

Pooter, Michiel De. 2007. "Examining the Nelson-Siegel class of term structure models." Tinbergen Institute Discussion Paper.

Prado, R., and M. West. 2010. *Time Series: Modeling, Computation, and Inference. Chapman & Hall/CRC Texts in Statistical Science*, Taylor & Francis.

Raftery, Adrian E., Miroslav Kárnỳ, and Pavel Ettler. 2010. "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill." *Technometrics*, 52(1): 52–66.

Rapach, David E., Jack K. Strauss, and Guofu Zhou. 2009. "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy." *Review of Financial Studies*, 23(2): 821–862.

Rossi, Barbara. 2005. "Optimal tests for nested model selection with underlying parameter instability." *Econometric theory*, 21(05): 962–990.

Sangvinatsos, Antonios, and Jessica A. Wachter. 2005. "Does the failure of the expectations hypothesis matter for long-term investors?" *The Journal of Finance*, 60(1): 179–230.

Sargent, Thomas J. 2002. *The Conquest of American Inflation.* Princeton University Press.

Sarno, Lucio, Daniel L. Thornton, and Giorgio Valente. 2007. "The empirical failure of the expectations hypothesis of the term structure of bond yields." *Journal of Financial and Quantitative Analysis*, 42(01): 81–100.

Sarno, Lucio, Paul Schneider, and Christian Wagner. 2014. "The economic value of predicting bond risk premia: Can anything beat the expectations hypothesis." Cass Business School and Centre for Economic Policy Research (CEPR) Working Paper.

Sharpe, William F. 1970. *Portfolio Theory and Capital Markets.* Vol. 217, McGraw-Hill New York.

Siegel, Andrew F., and Charles R. Nelson. 1988. "Long-term behavior of yield curves." *Journal of Financial and Quantitative Analysis*, 23(01): 105–110.

Stock, James H., and Mark W. Watson. 2002. "Forecasting using principal components from a large number of predictors." *Journal of the American Statistical Association*, 97(460): 1167–1179.

Tang, Huarong, and Yihong Xia. 2007. "An international examination of affine term structure models and the expectations hypothesis." *Journal of Financial and Quantitative Analysis*, 42(01): 41–80.

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

Timmermann, Allan G. 1993. "How learning in financial markets generates excess volatility and predictability in stock prices." *The Quarterly Journal of Economics*, 108(4): 1135–1145.

Timmermann, Allan G. 1996. “Excess volatility and predictability of stock prices in autoregressive dividend models with learning.” *The Review of Economic Studies*, 63(4): 523–557.

Ulrich, Maxim. 2013. “Inflation ambiguity and the term structure of us government bonds.” *Journal of Monetary Economics*, 60(2): 295–309.

Uppal, Raman, and Tan Wang. 2003. “Model misspecification and underdiversification.” *The Journal of Finance*, 58(6): 2465–2486.

Van Dijk, Dick, Siem Jan Koopman, Michel Van der Wel, and Jonathan H. Wright. 2014. “Forecasting interest rates with shifting endpoints.” *Journal of Applied Econometrics*, 29(5): 693–712.

Wan, Eric A., and Rudolph Van Der Merwe. 2000. “The unscented Kalman filter for nonlinear estimation.” In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*. 153–158. IEEE.

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

West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag New York, Inc.

Wheelock, David C., and Mark E. Wohar. 2009. “Can the term spread predict output growth and recessions? A survey of the literature.” *Federal Reserve Bank of St. Louis Review*, 91(Part 1): 419–440.

Wright, Jonathan H. 2006. “The yield curve and predicting recessions.” Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board Working Paper 07.

Wright, Jonathan H. 2011. “Term premia and inflation uncertainty: Empirical evidence from an international panel dataset.” *American Economic Review*, 101(4): 1514–1534.

Appendices

Data Appendix

Table 6: *List of Yields and Macro-Finance Variables*

Series ID	Description
TB	3- and 6-month Treasury Bills (Secondary Market Rate) [1]
ZCY	Smoothed Zero-coupon Yield from Gürkaynak, Sack and Wright (2007) [1]
IND	Industrial Production Index [5]
CPI	Consumer Price Index for All Urban Consumers: All Items Less Food & Energy [5]
FED	Effective Federal Funds Rate, End of Month [1]
SP	S&P 500 Stock Price Index, End of Month [5]
TCU	Capacity Utilization: Total Industry [1]
M1	M1 Money Stock [5]
TCC	Total Consumer Credit Owned and Securitized, Outstanding (End of Month) [5]
LL	Loans and Leases in Bank Credit, All Commercial Banks [5]
DOE	DOE Imported Crude Oil Refinery Acquisition Cost [5]
MSP	Median Sales Price for New Houses Sold in the United States [5]
TWX	Trade Weighted U.S. Dollar Index: Major Currencies [1]
ED	Eurodollar Spread: 3m Eurodollar Deposit Rate - 3m Treasury Bill Rate [1]
WIL	Wilshire 5000 Total Market Index [5]
DYS	Default Yield Spread: Moodys BAA-AAA [1]
NFCI	National Financial Conditions Index [1]

Notes:

1. In square brackets $[\cdot]$ we have a code for data transformations used in this data set: $[1]$ means original series is used; $[5]$ means log first-order difference is used to detrend and ensure stationarity. The series are seasonally adjusted when appropriate.
2. Data are obtained from St. Louis Federal Reserve Economic Data [http://research.stlouisfed.org/], spanning from Nov. 1971 to Nov. 2013. The smoothed zero-coupon yield is available on the Federal Reserve Board website [http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html/].
3. National Financial Conditions Index, provided by the Chicago Fed, is available on the website [http://www.chicagofed.org/webpages/publications/nfci/].
4. The small-size VAR model includes no macro variables. The medium-size VAR model includes only three macro variables: IND, CPI and FED. The large-size VAR model uses all the macro and financial variables in this data list.

A.1 Econometric Methods

A.1.1 Bayesian Kalman Filter with Forgetting Factor

We conduct the Kalman filter estimation for the state space model with Eq. (2.3) and Eq. (2.4):

$$z_t = X_t \beta_t + v_t,$$

$$\beta_{t+1} = \beta_t + \mu_t,$$

where z_t is an $n \times 1$ vector of variables, $X_t = I_n \otimes \left[z'_{t-1}, ..., z'_{t-p}\right]'$, β_t are VAR coefficients, $v_t \sim N(0, \Sigma_t)$ with Σ_t an $n \times n$ covariance matrix, and $\mu_t \sim N(0, Q_t)$.

Given that all the data from time 1 to t denoted as D_t, the Bayesian solution to updating about the coefficients β_t takes the form

$$\begin{aligned} p(\beta_t|D_t) &\propto \mathbf{L}(\beta_t; z_t)\, p(\beta_t|D_{t-1}), \\ p(\beta_t|D_{t-1}) &= \int_{\wp} p(\beta_t|D_{t-1}, \beta_{t-1})\, p(\beta_{t-1}|D_{t-1})\, d\beta_{t-1}, \end{aligned}$$

where $\wp$ is the support of β_{t-1}. The solution to this problem can be defined using a Bayesian generalization of the typical Kalman filter recursions. Given an initial condition $\beta_0 \sim N(m_0, \Phi_0)$ we can define (cf. West and Harrison (1997))[13]:

1. Posterior at time $t-1$

$$\beta_{t-1}|D_{t-1} \sim N(m_{t-1}, \Phi_{t-1}),$$

2. Prior at time t

$$\beta_t|D_{t-1} \sim N\left(m_{t|t-1}, \Phi_{t|t-1}\right),$$

   where $m_{t|t-1} = m_{t-1}$ and $\Phi_{t|t-1} = \Phi_{t-1} + Q_t$.

3. Posterior at time t

$$\beta_t|D_t \sim N(m_t, \Phi_t), \tag{13}$$

   where $m_t = m_{t|t-1} + \Phi_{t|t-1} X'_t (V_t^{-1})' \widetilde{v}_t$ and $\Phi_t = \Phi_{t|t-1} - \Phi_{t|t-1} X'_t (V_t^{-1})' X_t \Phi'_{t|t-1}$, with $\widetilde{v}_t = z_t - X_t m_{t|t-1}$ the prediction error and $V_t = X_t \Phi_{t|t-1} X'_t + \Sigma_t$ its covariance matrix.

Following the discussion above, we need to find estimates for Σ_t and Q_t in the formulas above. We define the time t prior for Σ_t to be

$$\Sigma_t|D_{t-1} \sim iW(S_{t-1}, \delta n_{t-1}), \tag{14}$$

[13] For a parameter θ we use the notation $\theta_{t|s}$ to denote the value of parameter θ_t given data up to time s (i.e. $D_{1:s}$) for $s > t$ or $s < t$. For the special case where $s = t$, I use the notation $\theta_{t|t} = \theta_t$

while the posterior takes the form

$$\Sigma_t | D_t \sim iW\left(S_t, n_t\right),$$

where $n_t = \delta n_{t-1} + 1$ and $S_t = \delta S_{t-1} + n_t^{-1}\left(S_{t-1}^{0.5} V_{t-1}^{-0.5} \widetilde{v}_{t|t-1} \widetilde{v}'_{t|t-1} V_{t-1}^{-0.5} S_{t-1}^{0.5}\right)$. In this formulation, v_t is replaced with the one-step ahead prediction error $\widetilde{v}_{t|t-1} = z_t - m_{t|t-1} X_t$. The estimate for Σ_t is approximately equivalent numerically to the Exponentially Weighted Moving Average (EWMA) filter $\hat{\Sigma}_t = \delta \hat{\Sigma}_{t-1} + (1-\delta) v_t v_t'$. The parameter δ is the decay factor, where for $0 < \delta < 1$. In fact, Koop and Korobilis (2013) apply such a scheme directly to the covariance matrix Σ_t, which results in a point estimate. In this case by applying variance discounting methods to the scale matrix S_t, we are able to approximate the full posterior distribution of Σ_t.

Regarding Q_t, we use the forgetting factor approach in Koop and Korobilis (2013); see also West and Harrison (1997) for a similar discounting approach. In this case Q_t is set to be proportionate to the filtered covariance $\Phi_{t-1} = cov\left(\beta_{t-1} | D_{t-1}\right)$ and takes the following form

$$Q_t = \left(\Lambda^{-1} - 1\right) \Phi_{t-1}, \tag{15}$$

for a given forgetting factor Λ.

The brief interpretation of forgetting factors is that they control how much 'recent past' information will be used. With the exponential decay for the forgetting factors, if it takes a value of 0.99, the information 24 periods ago (two years for monthly data) receives around 80% as much weight as the information of last period. If forgetting factor takes 0.95, then forecast performance 24 periods ago receives only about 30% as much weight. The similar implication holds for the decay factor.

A.1.2 Probabilities for Dynamic Selection and Averaging

To obtain the desire probabilities for dynamic selection or averaging, we need updating at each point in time. In papers such as Raftery, Kárnỳ and Ettler (2010) or Koop and Korobilis (2012) the models are TVP regressions with different sets of explanatory variables. The analogous result of the model prediction equation, when doing DMA or DPS, is

$$p(\beta_{t-1}|D_{t-1}) = \sum_{k=1}^{K} p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1})\mathbf{Pr}(L_{t-1} = k|D_{t-1}), \tag{16}$$

where $L_{t-1} = k$ means the k_{th} model[14] is selected and $p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1})$ is given by the Kalman filter (Eq. 53). To simplify notation, let $\pi_{t|s}^{(l)} = \mathbf{Pr}(L_t = l|D_s)$.

The model updating equation is

$$\pi_{t|t}^{(i)} = \frac{\pi_{t|t-1}^{(i)} p^{(i)}(z_t|D_{t-1})}{\sum_{l=1}^{K} \pi_{t|t-1}^{(l)} p^{(l)}(z_t|D_{t-1})}, \tag{17}$$

where $p^{(i)}(z_t|D_{t-1})$ is the predictive likelihood. Raftery, Kárnỳ and Ettler (2010) used an empirically sensible simplification that

$$\pi_{t|t-1}^{(i)} = \frac{\left(\pi_{t-1|t-1}^{(i)}\right)^{\alpha}}{\sum_{l=1}^{K} \left(\pi_{t-1|t-1}^{(l)}\right)^{\alpha}}, \tag{18}$$

where $0 < \alpha \leq 1$. A forgetting factor is also employed here, of which the meaning is discussed in the last section. The huge advantage of using the forgetting factor α is that it does not require an MCMC algorithm to draw transitions between models or a simulation algorithm over model space.

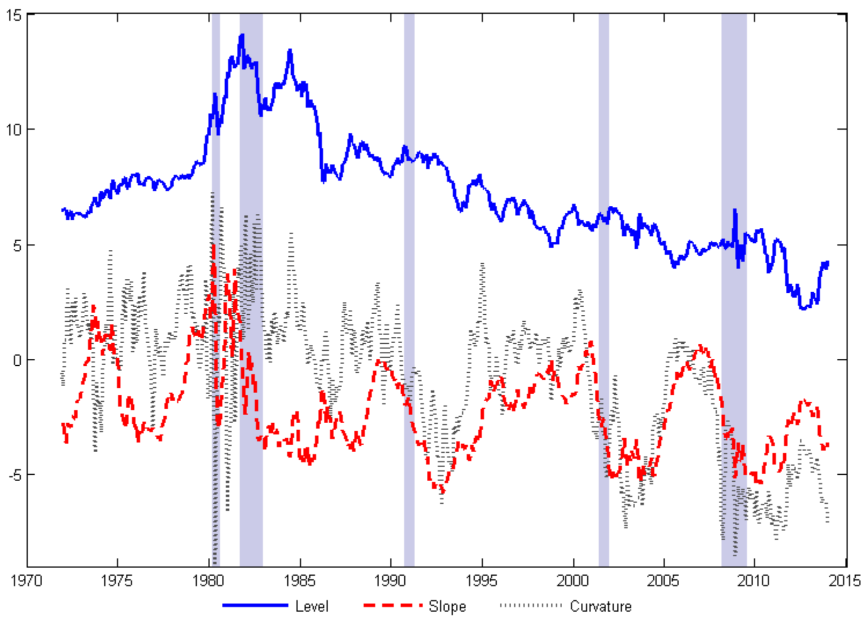
When proceeding with Dynamic Model Selection, the model with the highest probability is the best model we would like to select. Alternatively, we can conduct Dynamic Model Averaging, which average the predictions of all models with respective probabilities.

[14] For example, the k_{th} model in Dynamic Model Selection/Averaging, or the k_{th} candidate γ value in Dynamic Prior Selection.

A.2 Interpretation of Factor Dynamics

We illustrate the factor dynamics in this section and try to shed light on the economic implications of the latent factors. The extracted NS factors are shown in Figure 8. The Level factor has a downward trend since the early 1980s. The Level factor also has greater persistence compared with the other more volatile factors. The downward trend in the Level factor is consistent with the descriptive statistics in Table 3.1 and the results of Koopman, Mallee and Van der Wel (2010). The latter suggest a strong link between the Level factor and (expected) inflation, as they share high persistence. Evans and Marshall (2007) also indicate that there is a link between the level of yields and inflation with structural VAR evidence. In particular, the Level factor fell significantly after the financial crisis, which may indicate that the market had low inflation expectations. The Level factor rises in 2013, potentially associated with rising inflation and the impact of the Fed's Quantitative Easing (QE) pattern.

Figure 8: *Nelson-Siegel Factor Dynamics*



Notes: The graph shows the Nelson-Siegel Level, Slope and Curvature factors, which are drawn from Eq. (10). The shaded areas are recession periods according to the NBER Recession Indicators.

The Slope factor tends to fall sharply within recession periods, as indicated in Figure 8 by the shaded areas. Therefore, this factor could be closely related to real activity. The Slope factor is often considered as a proxy for the term spread, see Diebold, Rudebusch and Aruoba (2006). It can also be considered as a proxy for the stance of monetary policy, as the short end is influenced by policy rates.[15]

[15] Recent research relates the Slope of term structure to news shocks on total factor productivity and asset-class risk, see Kurmann and Otrok (2013) and Bansal, Connolly and Stivers (2014).

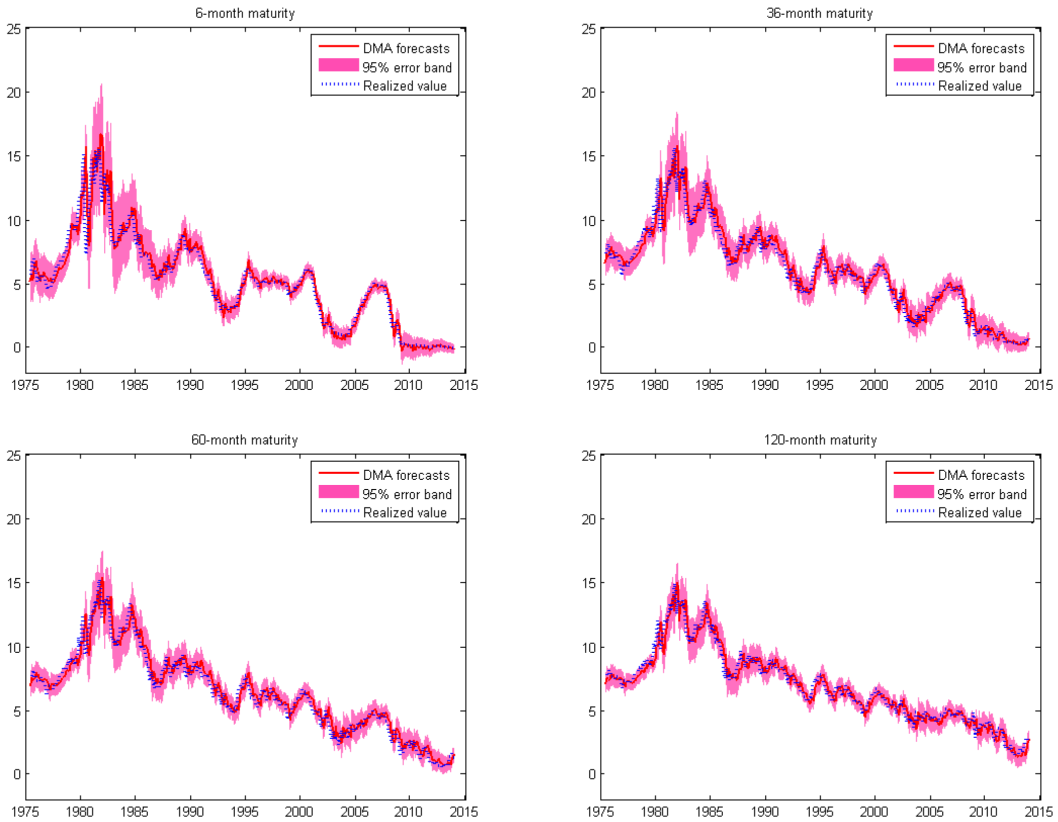
Lastly, the Curvature factor is harder to interpret and Diebold and Rudebusch (2013) indicate that this factor is less important than the other factors. On one hand, Litterman, Scheinkman and Weiss (1991) link the Curvature factor to the volatility of the level factor, via the argument of yield curve convexity, which can also be seen in Neftci (2004).[16] On the other hand, medium rates can be linked to expect short rates in the future, and therefore should be linked to current and expected future policies, which may potentially contain useful macro information missing in the first two factors.

[16]Generally, higher convexity means higher price-volatility or risk, and vice versa.

A.3 Additional Results

A.3.1 Forecasting Results

Figure 9: *DMA Forecasts of Yields*



Notes: These are 3 months ahead forecasts (95% error band) for yields against realized values with maturities 6, 36, 60 and 120 months, from early 1975 to late 2013. The forecasts are two-step forecasting using DMA, which can be summarized by Eq. (10), (2.3) and (2.4).

Table 7: *Relative MAFE Performance of Term Structure Models*

FH	h=1								h=3							
Maturity	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**
3	0.851†	0.890	0.866	0.858	0.968	1.050	0.979	1.275	0.884†	0.948	0.939	0.926	1.077	1.301	1.092	1.527
6	0.976†	1.036	1.025	1.007	1.101	1.218	1.120	1.462	0.957†	1.019	1.021	1.028	1.189	1.391	1.206	1.621
12	1.023†	1.047	1.021	1.039	0.984	0.963	1.013	1.324	0.987†	1.031	1.017	1.041	1.059	1.128	1.067	1.408
24	1.028†	1.053	1.041	1.047	1.047	1.059	1.054	1.272	1.006†	1.034	1.039	1.053	1.122	1.201	1.127	1.289
36	0.990†	1.010	1.011	1.015	1.062	1.142	1.073	1.131	1.013†	1.033	1.047	1.044	1.145	1.242	1.151	1.194
48	0.973†	0.987	1.000	0.992	1.065	1.164	1.072	1.049	1.012†	1.020	1.041	1.031	1.145	1.255	1.149	1.128
60	0.976†	0.986	1.000	0.988	1.061	1.167	1.066	1.023	1.011†	1.013	1.035	1.028	1.130	1.254	1.134	1.096
72	0.984†	0.995	1.006	1.000	1.051	1.160	1.059	1.030	1.008†	1.011	1.031	1.028	1.110	1.243	1.113	1.090
84	0.993†	1.006	1.014	1.013	1.041	1.140	1.047	1.050	1.003†	1.011	1.027	1.028	1.082	1.218	1.091	1.094
96	0.997†	1.014	1.020	1.025	1.030	1.110	1.038	1.071	0.998†	1.011	1.023	1.028	1.059	1.186	1.068	1.108
108	1.004†	1.023	1.027	1.037	1.015	1.069	1.029	1.091	0.994†	1.014	1.021	1.031	1.039	1.150	1.048	1.126
120	1.010†	1.028	1.034	1.044	1.004	1.028	1.019	1.110	0.988†	1.013	1.017	1.030	1.018	1.109	1.029	1.139
Mean	0.988†	1.008	1.008	1.009	1.036	1.104	1.047	1.143	0.991†	1.015	1.023	1.027	1.098	1.220	1.106	1.224

Notes: 1. This table shows 1-month and 3-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from 1983:10 to 2013:11.
2. The MAFE-based statistics relative to the RW are reported. The dagger (†) indicates, in terms of the sum of predictive log-likelihood, the model has the preferred value among all models at certain maturities (or in total), see Geweke and Amisano (2010).
3. In this table, we use following abbreviations. **MAFE**: Mean Absolute Forecasting Error; **Mean**: Averaged MSFE across all sample maturities. **DMA** (Dynamic Model Averaging) averages all the models with probabilities in each step, while **DMS** (Dynamic Model Selection) chooses the best model with the highest probability at any point in time. **TVP-M**: a time-varying parameter model with three macro variables: fund rate, inflation and industrial production, similar to Bianchi Mumtaz and Surico (2009) but estimated with a fast algorithm without the need of MCMC; **DL**: Diebold and Li (2006) model, i.e. constant coefficient Vector Autoregressive model with recursive (expanding) estimations; **DL-R10**: Diebold and Li (2006) estimates based 10-year rolling windows; **TVP**: a time-varying parameter model without macro information; **DL-M**: factor dynamics in Diebold and Li (2006) are augmented with three macro variables: fund rate, inflation and industrial production, using recursive estimations; **DL-SW**: factor dynamics in Diebold and Li (2006) are augmented with three principal components (see Stock and Watson (2002)) of our macro-finance data, using recursive estimations; **RW**: Random Walk.

Table 8: *Relative MAFE Performance of Term Structure Models (Continued)*

FH	h=6								h=12							
Maturity	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**	**DMA**	**DMS**	**TVP**	**TVPM**	**DL**	**DLR10**	**DLM**	**DLSW**
3	0.951†	1.009	1.020	1.003	1.206	1.360	1.249	1.522	1.028†	1.074	1.078	1.127	1.220	1.321	1.296	1.374
6	0.999†	1.055	1.068	1.061	1.250	1.411	1.290	1.581	1.054†	1.098	1.101	1.141	1.245	1.365	1.313	1.438
12	0.995†	1.052	1.044	1.047	1.131	1.236	1.148	1.445	1.031†	1.085	1.073	1.082	1.190	1.308	1.229	1.389
24	0.999†	1.040	1.036	1.033	1.165	1.263	1.176	1.348	1.055†	1.095	1.093	1.068	1.242	1.392	1.274	1.363
36	1.017†	1.040	1.046	1.027	1.187	1.297	1.197	1.290	1.063†	1.092	1.097	1.054	1.258	1.433	1.287	1.313
48	1.022†	1.036	1.046	1.024	1.187	1.305	1.194	1.243	1.061†	1.086	1.089	1.046	1.259	1.448	1.287	1.291
60	1.017†	1.029	1.040	1.021	1.174	1.296	1.177	1.212	1.054†	1.079	1.080	1.039	1.250	1.448	1.276	1.287
72	1.008†	1.022	1.033	1.019	1.153	1.277	1.155	1.199	1.048†	1.073	1.070	1.034	1.239	1.442	1.266	1.295
84	0.999†	1.015	1.025	1.016	1.127	1.250	1.128	1.196	1.041†	1.069	1.062	1.029	1.227	1.431	1.254	1.313
96	0.993†	1.011	1.020	1.014	1.102	1.227	1.104	1.205	1.032†	1.066	1.054	1.025	1.212	1.415	1.239	1.336
108	0.985†	1.008	1.015	1.012	1.079	1.205	1.080	1.215	1.019†	1.061	1.045	1.020	1.193	1.393	1.219	1.357
120	0.979†	1.007	1.012	1.012	1.061	1.183	1.061	1.228	1.007†	1.056	1.038	1.016	1.171	1.369	1.198	1.378
Mean	0.998†	1.028	1.034	1.025	1.153	1.276	1.164	1.306	1.042†	1.079	1.075	1.060	1.227	1.395	1.264	1.346

Notes: 1. This table shows six-month and twelve-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from early 1983 to the end of 2013.
2. The MAFE-based statistics relative to the RW are reported. The dagger (†) indicates, in terms of the sum of predictive log-likelihood, the model has the preferred value among all models at certain maturities (or in total), see Geweke and Amisano (2010).
3. In this table, we use following abbreviations. **MAFE**: Mean Absolute Forecasting Error; **Mean**: Averaged MAFE across all sample maturities. **DMA** (Dynamic Model Averaging) averages all the models with probabilities in each step, while **DMS** (Dynamic Model Selection) chooses the best model with the highest probability at any point in time. **TVP-M**: a time-varying parameter model with three macro variables: fund rate, inflation and industrial production, similar to Bianchi Mumtaz and Surico (2009) but estimated with a fast algorithm without the need of MCMC; **DL**: Diebold and Li (2006) model, i.e. constant coefficient Vector Autoregressive model with recursive (expanding) estimations; **DL-R10**: Diebold and Li (2006) estimates based 10-year rolling windows; **TVP**: a time-varying parameter model without macro information; **DL-M**: factor dynamics in Diebold and Li (2006) are augmented with three macro variables: fund rate, inflation and industrial production, using recursive estimations; **DL-SW**: factor dynamics in Diebold and Li (2006) are augmented with three principal components (see Stock and Watson (2002)) of our macro-finance data, using recursive estimations; **RW**: Random Walk.

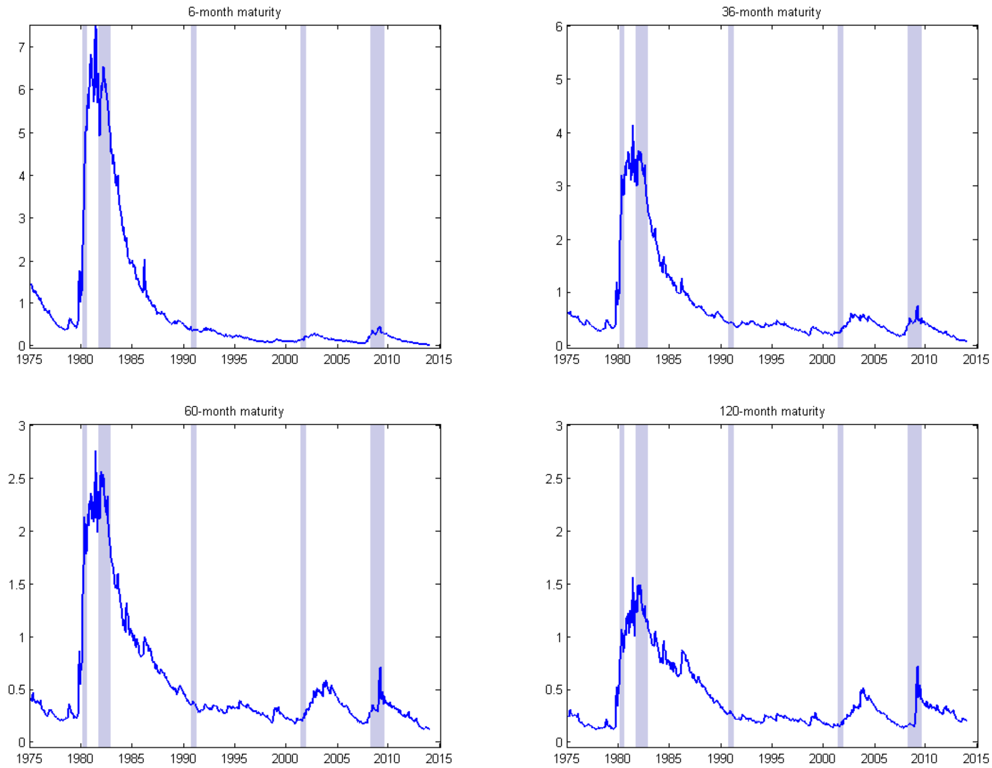
A.3.2 Time-Varying Volatility

It has been indicated by Bianchi, Mumtaz and Surico (2009) that homoskedasticity is a frequent and potentially inappropriate assumption in much of the macro-finance literature. Cieslak and Povala (2015*b*) show that stochastic volatility can have a non-trivial influence on the conditional distribution of interest rates. Piazzesi (2010) indicates that fat tails in the distribution of bond factors can be modeled by specifying an appropriate time-varying volatility. The DMA model allows for heteroskedastic variances and this assumption is crucial for its good density forecast performance; this evidence is consistent with Hautsch and Yang (2012).

The DMA not only provides more sensible results in terms of density forecasts, but also captures the desirable evolutionary dynamics of the economic structure. Figure 10 shows the time-varying second moments of 3 month ahead forecasts from the DMA model. The figure displays distinct time variation in the evolution of volatility. The stable decline of volatility before the financial crisis matches the conclusions of Bianchi, Mumtaz and Surico (2009), who refer to this empirical result as the 'Great Moderation' of the term structure. We observe that yields with longer maturities have lower volatilities. This feature is counter-intuitive. Theoretically, long yields are mainly driven by three components: the expected future (real) short yields; inflation expectations; and the term premia. Inflation expectations may change abruptly and frequently during a short period of time, so do the expected future short yields. At the same time, term premia can also be quite volatile. Therefore, summing up the movements of these three components, the variance of long yields should be larger than the short yields; nevertheless, the empirical result implies the opposite. As indicated in Duffee (2011*b*), the reason causing this result is that the factor driving up the expected future short yields or inflation expectations may drive down the term premia, thus, offsetting the variation in these components.

From the perspective of time dimension, the volatilities of yields (especially shorter-term) are high in the 1980s, while the bond yield level is also relatively high. The high volatilities are due to large forecast variances of forecast models as well as a high degree of forecast dispersion in forecasts. It is clear that the volatilities are declining during the Great Moderation, and therefore the variances of bond forecasts are rather small between 1990 and 2007, except during the 2004-05 episode of 'Greenspan's Conundrum'. In around 2009, the volatilities surge to a high level since the 1990's, although the short yields stay at a relatively low level (restricted by the zero lower bound) among all periods. Even after the financial crisis, ambiguity in yield forecasts still exists as the volatilities remain at a relatively high level.

Figure 10: *Time-Varying Second Moment*



Notes: These are time-varying second moments of 3 months ahead forecasts for bonds at maturities 6, 36, 60 and 120 months, from early 1975 to late 2013. The variance of NS factors is estimated from Eq. (54), and then the variances of yield forecasts generated by each candidate model in the DMA, can be easily calculated as linear combinations of factor variances.

A.3.3 Robustness: Do We Need Strict Arbitrage-Free Restrictions?

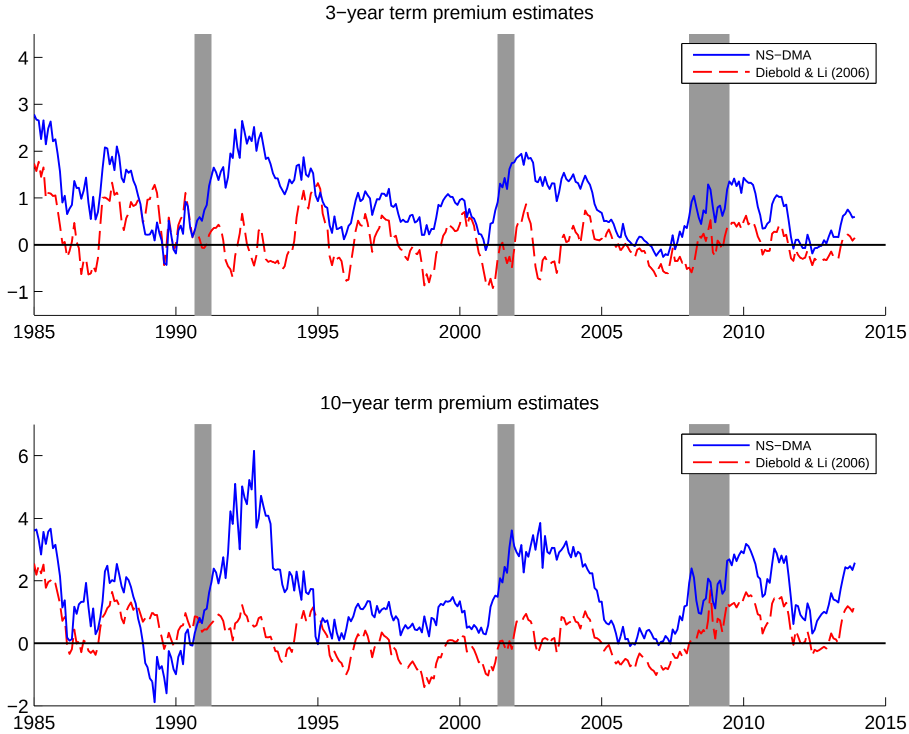
As we have discussed in Section 2.2, we impose NS restrictions on the pricing dynamics and leave the physical dynamics unconstrained. By allowing for parameter and model uncertainty in the physical dynamics, we are able to acquire significant predictive gains. The sources of these gains are also revealed in the last section.

Our DMA approach does not explicitly impose 'hard' arbitrage-free restrictions. From a theoretical perspective, Filipović (1999) and Björk and Christensen (1999) show that the Nelson-Siegel family does not impose the restrictions necessary to eliminate opportunities for riskless arbitrage. From a practical perspective, our implementation allows all bond yields to be priced with errors, which naturally breaks their original assumptions of the Nelson-Siegel family in their papers. Therefore, the potential loss of not imposing arbitrage-free restrictions may be mitigated. The reason is that our focus here is not on the dynamic structure of market price of risks. Duffee (2014) indicates that the no-arbitrage restrictions are unimportant, if a model aims to pin down physical dynamics but not equivalent-martingale dynamics that specify the pricing of risk. In order to capture expectations of investors, we aim to improve forecasts of the interest rate term structure. Joslin, Singleton and Zhu (2011) show that no-arbitrage cross-sectional restrictions are irrelevant for out-of-sample forecasts if the factor dynamics are unrestricted. In practice, the arbitrage-free restrictions are not important in terms of forecasting in models assuming bond yields are priced with errors, see for example, Coroneo, Nyholm and Vidova-Koleva (2011) and Carriero and Giacomini (2011).

To ensure the robustness of our DMA approach, we extend the three-factor arbitrage-free Nelson-Siegel model proposed by Christensen, Diebold and Rudebusch (2011) and evaluate the forecast performance of the arbitrage-free version of DMA. The key difference between arbitrage-free DMA and DMA is a 'yield-adjustment term', which only depends on the maturity and factor volatility. See Christensen, Diebold and Rudebusch (2011) and Diebold and Rudebusch (2013) for more details. The forecast performances of two models are very close, which implies that the DMA is almost arbitrage-free, which is consistent with theoretical evidence in Feunou et al. (2014) and Krippner (2015) that the NS models are near arbitrage-free. Hence, following Duffee (2014), we choose not to impose arbitrage-free restrictions to avoid potential misspecification.

A.3.4 Term Premia of Diebold-Li and DMA

Figure 11: *Time-Varying Term Premia of 36-and 120-Month Bonds*



Notes:
1. The top panel is the 36-month term premia and the bottom is the 120-month term premia. The EH consistent 36- and 120-month bond yields are estimated using Eq. (4.38); we then calculate the term premia using Eq. (4.41).
2. In addition to DMA, we plot the recursively estimated term premia employing the methods proposed by Diebold and Li (2006).
3. Shaded areas are recession periods based on the NBER Recession Indicators. The unit is percentage.

CHAPTER 3

Co-Movement and Spillovers in Global Bond Markets

ABSTRACT

This paper investigates global term structure dynamics using a Bayesian hierarchical factor model augmented with macroeconomic fundamentals. More than half of the variation in the bond yields of seven advanced economies is due to global co-movement. Our results suggest that global inflation is the most important factor among global macro fundamentals. Non-fundamental factors are essential in driving global co-movements, and are closely related to sentiment and economic uncertainty. Lastly, we analyze asymmetric spillovers in global bond markets connected to diverging monetary policies.

Keywords: Global Bond Markets, Term Structure of Interest Rates, Shocks to Fundamentals and Non-Fundamentals, Co-Movement, Contagion, Sentiment, Economic Uncertainty.
JEL Classification Codes: C11; C32; E43; F3; G12; G15.

Author Contributions: This chapter is drawn from the collaborative work with my supervisors Joseph P. Byrne and Dimitris Korobilis, and a working paper version is available online. I undertook the econometric analysis and did the vast majority of the writing.

3.1 Introduction

Reduced-form factor models are widely used in analyzing the term structure of interest rates. These factor models assume the yield curve is driven by a few pricing factors and can be divided into two groups. The first group of models, denoted 'fundamentals-driven', uses macroeconomic fundamentals as pricing factors and helps explain how these factors affect asset prices.[1] In contrast, term structure models using latent pricing factors have a more successful empirical fit and avoid the mispricing indicated by Anh and Joslin (2013). However, Diebold, Rudebusch and Aruoba (2006) suggest that latent factors are not explicitly linked to macroeconomic variables. The recent work of Joslin, Priebsch and Singleton (2014) and Bauer and Rudebusch (2015) reconciles the above seemingly contradictory evidence by proposing hybrid models that incorporate the joint dynamics of both fundamentals and latent factors.[2] Their work paves a way to understand the linkage between latent information and (non)fundamentals that have economic content.

Non-fundamentals are also essential to asset prices. For instance, Lee (1998) finds that only 10% of the variance of stock prices is driven by stock fundamentals. There have been various theories proposed to explore shocks to non-fundamentals. Sentiment is a popular explanation, and Bansal and Shaliastovich (2010) show that the variance of returns is more susceptible to a sentiment or confidence measure than fundamentals in the economy. In contrast, Bloom (2014) stresses the importance of economic uncertainty which affects agents' behavior and, therefore, asset prices. Novy-Marx (2014) reviews the earlier literature on non-fundamentals and suggests other potential explanations. While most of the current empirical research focuses on the driving forces of domestic asset prices, only a few studies try to approach this topic from a global perspective. Hou, Karolyi and Kho (2011) indicate that liberalization of financial markets around the world has increased market co-movement, but to what extent the co-movement is driven by global macroeconomic factors or non-fundamentals remains unanswered.

This paper aims to study the underlying drivers of global term structures by explicitly considering shocks to fundamentals and non-fundamentals. The contribution of Diebold, Li and Yue (2008) provides empirical evidence of strong co-movement in yield curves across countries. Den Haan and Sumner (2004) reveal global co-movements in real activity and prices, whereas Eickmeier, Gambacorta and Hofmann (2014) present evidence of global liquidity. One question is naturally raised; whether the co-movement in bond yields is determined by global fundamentals. We specifically tackle this question in a global context, as Barberis, Shleifer and Wurgler (2005) suggest that common movement of asset prices among international markets may not be easily explained by a fundamentals-based view. Our particular interest is twofold: How much of the variance in global bond yield co-movement is driven by global fundamentals, and why would that be the case? To answer the first question, we identify structural shocks of global fundamentals. To further

[1] See for example, Kozicki and Tinsley (2001), Dewachter and Lyrio (2008) and Orphanides and Wei (2012).

[2] Despite similar findings, their models differ in whether or not macro factors are fully spanned by bond yields.

understand the underlying mechanism, we decompose long yield movements into two transmission channels, viz., a 'policy channel' and a 'risk compensation channel'. These two standard channels are associated with short rate expectations and risk premia, respectively. We then evaluate the effects of global fundamentals through each channel and find this practice is informative in terms of economic implications.

This paper's main finding is in support of the sentiment-based theory favored by Kumar and Lee (2006), Bansal and Shaliastovich (2010) and Benhabib and Wang (2015), and the effects of economic uncertainty suggested by Bloom (2014). We pin down that more than 70% of bond yield co-movement is driven by shocks to non-fundamentals, and non-fundamental movements can be largely explained by the measures of investor sentiment and economic uncertainty. Among all fundamentals, global inflation has demonstrable influence on the co-movement of global short rates. With regard to the co-movements of long rates, the shocks to fundamentals are less significant.[3] This empirical evidence shows that the majority of variability in bond yield data can be satisfactorily captured by the information of global macro, sentiment and economic uncertainty, and, therefore, latent information has economic appeal. This evidence complements the study of Piazzesi and Schneider (2007) and calls for a more complete structural model with these salient features, and in particular, the consideration of global transmission mechanisms of sentiment and economic uncertainty.

To model the global term structures of seven advanced economies, we adopt the novel methodology of Moench, Ng and Potter (2013) and propose an augmentation with global macro variables. The three-hierarchy structure is a straightforward specification. At the highest global level, we allow global macroeconomic fundamentals to interact with global bond factors. At a lower level, national bond factors are driven by global bond factors and country-specific components. At the lowest level, the term structure of each country is driven by national bond factors and innovations. We introduce macro factors following the setup of Coroneo, Giannone and Modugno (2015) and Joslin, Priebsch and Singleton (2014) for parsimony. With the proposed model specification, we jointly identify global and national bond pricing factors in a one-step Bayesian approach. We find that two global yield factors can explain, on average, more than 60% of bond yields' variance across seven countries, and country-specific components contribute to most of the remaining variance. By conducting an analysis on country-specific components we duly unfold asymmetric spillovers among seven countries, which are linked to diverging monetary policies (Jotikasthira, Le and Lundblad (2015)).

Our work is related to the literature of global term structures. Bauer and Diez de los Rios (2012) model the unspanned macroeconomic risks driving international term premia and foreign exchange risk premia. In a similar framework but without international finance restrictions, Abbritti et al. (2013) reveal contrasting forces driving long- and short-term dynamics in yield curves. The most recent work of Jotikasthira, Le and Lundblad (2015) investigates three countries' bond yield co-movement before the financial crisis. The cross-sectional restrictions used are the same as Diebold, Li and

[3] Duffee (2011*b*) and Joslin, Priebsch and Singleton (2014) suggest shocks to fundamentals can be offset from one another, i.e. the shocks driving up expected future short yields drive down term premiums, which makes these factors unspanned or weakly identified.

Yue (2008), and the proposed extension allows a direct assessment of the internal link between co-movement in bond yields and shocks to fundamentals and non-fundamentals. Building upon previous work, we employ a parsimonious specification to jointly identify latent factors with the help of global fundamentals. This specification is new to the literature, as it considers both cross-sectional and time-series properties of the data, and thus the identified dynamic factors describe contemporaneous and temporal covariation among the variables. Specifically, the global dynamics are stably pinned down by a low dimension of latent factors, and the results are robust to the macro spanning condition.

The structure of the paper is as follows. In Section 3.2 we introduce the model and describe the estimation and identification of the model. In Section 3.3 we describe the data and present a preliminary data analysis. In Section 3.4 we report empirical results. In particular, we decompose the yield co-movements into two channels and distinguish the role of global inflation. Moreover, we find that non-fundamental factors, which are important in driving global co-movements, are closely related to sentiment and economic uncertainty. Section 3.4.4 sets out the asymmetric 'spillovers' in global bond markets. In Section 3.5 we perform robustness checks by testing whether the results are sensitive to the macro spanning condition and zero lower bound. In Section 3.6 we conclude and summarize the implications of this analysis.

3.2 Methodology

3.2.1 Model Specification

To analyze global bond yields, a hierarchical factor methodology is needed. The framework shall model bond yields across countries, using global macro and yield factors. In the spirit of multi-level factor models, Kose, Otrok and Whiteman (2003) propose a dynamic factor model to study international business cycle co-movements, whereas Moench, Ng and Potter (2013) propose a hierarchical factor model to study the US housing market.[4] In our specification, a hierarchical factor structure is relatively more parsimonious in terms of parameters to be identified and retains a low-dimensional factor structure, making it attractive for bond yield modeling. Building upon Kose, Otrok and Whiteman (2003) and Moench, Ng and Potter (2013), we allow the dynamic factors to interact with each other at the global level.

The parsimony is also related to the setup of Joslin, Priebsch and Singleton (2014) and Coroneo, Giannone and Modugno (2015), who impose knife-edge restrictions on the loadings of bond yields so that the macro factors cannot be inverted from yields. They denote this setting as *Unspanned Macro Risks* and argue that it is a more realistic assumption. By definition, if there exist Unspanned Macro Risks, macro factors do not directly or contemporaneously impact yields and they influence current yields only through their correlation with the yield factors. Joslin, Priebsch and Singleton (2014) suggest that the fully spanned assumption, i.e. the macro factors can be inverted as linear combinations of yields, is often questioned and might be counterfactual.[5] We will show later the macroeconomic variables still have significant effects on bond yields movements without forcing macro factors to be a linear combination of bond yields.

The model for bond yields X_{ibt} can be written as:

$$X_{ibt} = \Lambda^F_{ib} F_{bt} + e^X_{ibt}, \tag{3.1}$$

$$F_{bt} = \Lambda^G_b G_t + e^F_{bt}, \tag{3.2}$$

$$\begin{bmatrix} G_t \\ M_t \end{bmatrix} = \psi^G \begin{bmatrix} G_{t-1} \\ M_{t-1} \end{bmatrix} + u_t, \tag{3.3}$$

in which the subscript i indicates the maturities of bond yields, the subscript b indicates the countries and the subscript t indicates different periods of time. In the above, Λ^F_{ib}, Λ^G_b and ψ^G are model parameters, and e^X_{ibt}, e^F_{bt} and u_t are error terms. Note that each element in e^X_{ibt} and e^F_{bt} follows a first order autoregressive process, but we do not assume homoskedastic innovations for these error terms;

[4]Kose, Otrok and Whiteman (2003) identify regional factors that are uncorrelated with the global factors, while Moench, Ng and Potter (2013) aim to find the global factors driving the regional factors. In fact, two frameworks are compatible and Moench, Ng and Potter (2013) can be considered nested in Kose, Otrok and Whiteman (2003).

[5]We encourage interested readers to read Appendix B.1.1 for the economic implications of the macro spanning condition. We test the robustness of our specification in Section 3.5.

the covariance matrix of u_t is unrestricted. In the country-level Equation (3.1), X_{ibt} represent the bond yield of country b at maturity i, and F_{bt} are the latent yield factors of country b. In Equation (3.2), G_t are the latent global yield factors that drive the national yield factors F_{bt}. Finally Equation (3.3) describes the interactions between the yield factors and the global macro factors/fundamentals M_t using a Vector Autoregression (VAR).[6]

As there is no consensus about non-fundamentals in the current literature, we only include macro information to facilitate the identification, but we will explore the economic content of non-fundamental movements afterwards. In our model, we include four global macro variables: monetary policy rate, inflation, real activity and financial conditions, such that M_t is a 4×1 vector. The former three are standard macro fundamentals in term structure modeling, see for example, Ang and Piazzesi (2003). Additionally, we include financial conditions because liquidity and credit risk measures are suggested by Dewachter and Iania (2012).

A key feature of our model is to augment the VAR system of global yield factors with global macro factors M_t. By extending the 'Dynamic Hierarchical Factor Model' proposed by Moench, Ng and Potter (2013), the proposed model captures the interdependencies among global macro variables and pricing factors. The dynamics of the global factors are characterized by an unrestricted Factor Augmented Vector Autoregressive (FAVAR) model. Factor augmentation has various advantages as suggested by Bernanke and Boivin (2003), and it is also of importance in the context of this paper. Global macro factors are incorporated to provide an economic interpretation of yield movements and exploit the underlying dynamics. Moreover, incorporating the information drawn from a large set of variables is helpful to negate the potential non-fundamentalness of the VAR, as suggested by Fernández-Villaverde and Rubio-Ramírez (2007) and Leeper, Walker and Yang (2013). The extended version of the hierarchical model is denoted as 'Fundamentals-Augmented Hierarchical Factor Model' (FAHFaM).[7]

The model proposed in this paper has a similar structure to Diebold, Li and Yue (2008) but contrasts in that their approach uses two steps and does not include macro determinants. We adopt a one-step Bayesian technique which should provide more accurate estimates. Diebold, Rudebusch and Aruoba (2006) and Pooter (2007) provide evidence that a one-step approach produces more effective estimates: Two-step estimation introduces bias if it does not fully consider the dynamics of the factors at a higher level. As shown in the previous literature, directly introducing macro fundamentals can provide a meaningful narrative which delineates the macro shocks that drive global term structures. Our hierarchical one-step framework allows us to jointly estimate the global factors and country-specific components, and hence builds upon the contribution of Bauer and Diez de los Rios (2012), Abbritti et al. (2013) and Jotikasthira, Le and Lundblad (2015). Identification schemes of structural shocks can be directly introduced in this one-step approach and posterior coverage intervals are readably available, without running additional regressions that can potentially introduce bias.

[6]When referring to global macro fundamentals, 'fundamental' and 'factor' are used interchangeably in this paper.

[7]Technical details of our FAHFaM are summarized in Appendix B.2.

3.2.2 Identification

To identify the global factors, a standard approach is the Principal Component method, but this lacks the consideration of the underlying time-series dynamics and hence may not be plausible in revealing the structural shocks. In this paper therefore we use an alternative identification scheme to Moench, Ng and Potter (2013) and impose cross-sectional restrictions. While these authors use zero restrictions which are of a statistical nature, we impose restrictions implied by the dynamic Nelson-Siegel (NS) term-structure model. In other words, the loadings of country-level factors are exactly the same as in Diebold, Li and Yue (2008).[8] The NS identification scheme has gained great popularity in term structure modeling, and we choose this scheme to fix the ideas.

We closely follow Diebold, Li and Yue (2008) to impose cross-sectional restrictions and only specify two global factors, as Bauer and Hamilton (2015) suggest that only the Level and the Slope factors are robust predictors of excess bond returns. This seems in contrast with the studies of Moench (2012) and Abbritti et al. (2013) that an additional factor (Curvature) is helpful in revealing the term premium dynamics. Indeed, without macro information, the term premium dynamics in our sample varies substantially if the Curvature factor is added. However, with a global macro augmentation, the term premium dynamics are not sensitive the number of factors and we are able to use this more parsimonious parameterization.[9] This is due to the nature of our identification strategy: The identified factors incorporate the time-series information of weakly identified factors, which also emphasizes the novelty of our global macro augmentation.

We estimate our model using a Bayesian estimation technique, specifically, the Gibbs sampling. Following Moench, Ng and Potter (2013), we assume that the prior distribution for all factor loadings coefficients is Gaussian, and the prior distribution for the variance parameters is a scaled inverse chi-square distribution.[10] These conjugate priors simplify the estimation problem, both mathematically and computationally. For the FAVAR of global dynamics we use uninformative priors, see Koop and Korobilis (2009). In the Gibbs sampling, we begin with $50,000$ burn in draws and then save every 50th of the remaining $50,000$ draws. These 1000 draws are used to compute posterior means and standard deviations of the factors, as well as the posterior coverage intervals in the following sections.

We identify global structural shocks using Cholesky decomposition, see Appendix B.2.3. The ordering of our global VAR system is the following: Industrial Production growth rate, inflation rate, change in policy rate, Level, Slope and FCI. The ordering of the first three variables is standard in the related literature, for example Christiano, Eichenbaum and Evans (2005). These three are

[8] The details of the restrictions can be found in Appendix B.2.2. The two schemes, Diebold, Li and Yue (2008) and Moench, Ng and Potter (2013), share similar results, as shown in Appendix B.4.1. In fact, the identified factors from two schemes are nearly identical subject to rotations. For more information regarding factor identification we refer the reader to Bai and Wang (2015).

[9] The results are qualitatively and quantitatively similar with a resonantly small number of factors (≤ 5), and are available upon request.

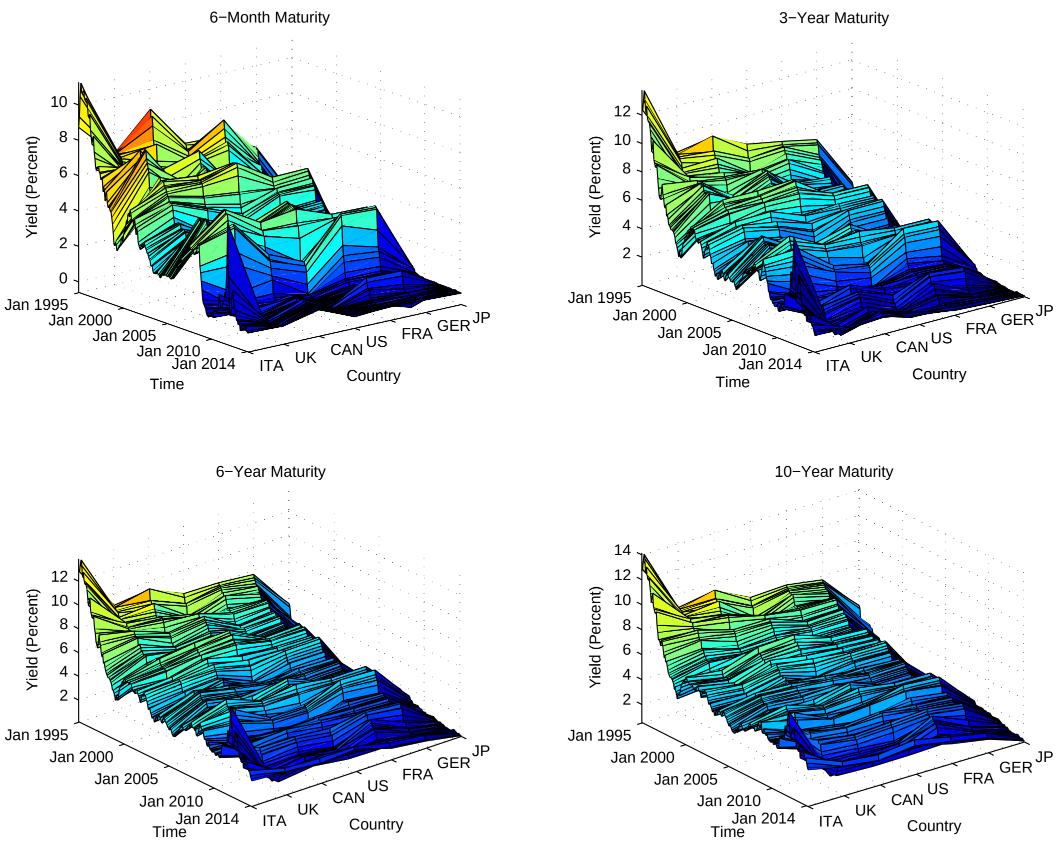
[10] The specified prior distributions are $N(0,1)$ and Scale-inv-$\chi^2(4,0.1^2)$ for loading and variance parameters, respectively.

followed by the financial variables Level, Slope and FCI, so the financial variables can react to the contemporaneous macro shocks in the first three variables. The Level, Slope and FCI are placed lower in the ordering because Hubrich, D'Agostino et al. (2013) argue that the monetary policy only react to asset price movements if there are prolonged, while the bond yields react immediately to policy change. It is worth noting that we do not find a significant difference for macro shocks using alternative orderings.

3.3 Data Description and Preliminary Evidence

We obtain monthly bond yield data from Bloomberg for seven advanced countries: Italy, Canada, France, Germany, Japan, the UK and the US. The empirical analysis focuses on government yields of 11 maturities: 3, 6, 12, 24, 36, 48, 60, 72, 84, 96 and 120 months. Figure 3.1 shows the dynamics of bond yield at four maturities across all seven countries. All four maturities trend down from the beginning of the sample period, with the shorter rates displaying greater variance across time and countries.

Figure 3.1: *Bond Yields of Seven Countries*



Notes:
1. The above charts plot the bond yields for the seven countries in the sample. The sample includes Italy (ITA), Canada (CAN), France (FRA), Germany (GER), Japan (JP), the UK and the US, spanning from Dec. 1994 to Mar. 2014.
2. From top left clock-wise we have bond yields of maturities 6 months, 3 year, 10 years and 6 years. More information about the data is provided in Appendix B.3.

Our empirical model uses macroeconomic variables from Bloomberg, and indicators of financial condition from St. Louis Federal Reserve Economic Data (FRED). We construct four global macro factors using a list of macro fundamentals among the seven countries, and the fundamentals include inflation (CPI), Industrial Production (IP) and the change in monetary policy rates. We also use a large number of regional series of Financial Condition Index (FCI) to construct a global FCI. The full sample of monthly data is from December 1994 to March 2014. The details about the data are described in the Data Appendix B.3.

Before we implement our one-step estimation, the global macro factors M_t are extracted from regional macro series. There are four categories of regional macro series: the change in policy rate, indicator of real activity, inflation and Financial Condition Index (FCI). We employ a new approach proposed by Koop and Korobilis (2014) to extract the global macro indicators from regional series.[11] Figure 8 in Appendix B.4.2 displays the estimated macro factors used to augment our proposed model.

In Table 3.1 we report summary statistics for bond yields at representative maturities. All yield curves are upward-sloping, suggesting positive term spreads. Apart from Japan, the yield volatility generally decreases with maturity, and all the yields are highly persistent for all countries, with first-order autocorrelation greater than 0.95. Japanese yields are typically the lowest, usually below two percent and are less persistent when compared to other yields.

Table 3.1: *Descriptive Statistics of Bond Yields*

Country	Maturity	Mean	Std. Dev.	Min.	Max.	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
US	3	2.82	2.27	0.01	6.39	0.99	0.74	0.26
	12	3.04	2.26	0.10	7.20	0.98	0.76	0.30
	60	3.92	1.88	0.59	8.03	0.97	0.76	0.42
	120	4.57	1.45	1.54	8.00	0.97	0.72	0.43
UK	3	3.91	2.32	0.28	7.50	0.99	0.77	0.47
	12	4.00	2.36	0.12	7.45	0.99	0.77	0.48
	60	4.51	2.00	0.58	8.94	0.98	0.75	0.43
	120	4.85	1.66	1.57	8.90	0.97	0.72	0.29
Germany	3	2.49	1.52	0.00	5.14	0.98	0.66	0.27
	12	2.63	1.53	0.01	5.82	0.98	0.64	0.25
	60	3.48	1.55	0.33	7.47	0.97	0.67	0.35
	120	4.17	1.46	1.22	7.69	0.97	0.70	0.35
France	3	2.63	1.68	0.01	7.93	0.98	0.56	0.21
	12	2.76	1.66	0.02	7.04	0.97	0.58	0.22
	60	3.67	1.49	0.69	7.87	0.96	0.61	0.29
	120	4.42	1.32	1.85	8.14	0.96	0.63	0.30
Italy	3	3.44	2.58	0.28	11.00	0.98	0.63	0.24
	12	3.73	2.50	0.60	11.74	0.98	0.57	0.17
	60	4.90	2.39	1.95	14.01	0.96	0.51	0.11
	120	5.61	2.22	3.42	14.14	0.97	0.54	0.09
Canada	3	3.10	1.91	0.21	8.88	0.96	0.59	0.28
	12	3.36	1.90	0.49	8.88	0.97	0.64	0.33
	60	4.23	1.80	1.19	9.40	0.97	0.74	0.45
	120	4.75	1.69	1.72	9.48	0.97	0.74	0.41
Japan	3	0.25	0.34	0.00	2.24	0.89	0.28	0.07
	12	0.31	0.37	0.01	2.48	0.89	0.39	0.07
	60	0.91	0.66	0.13	4.07	0.92	0.57	0.17
	120	1.66	0.77	0.55	4.79	0.94	0.60	0.18

Notes: This table presents descriptive statistics for monthly yields at different maturities across seven countries. The sample period is 1994:12–2014:03. We use the following abbreviations. **Std. Dev.**: Standard Deviation; **Min.**: Minimum; **Max.**: Maximum; $\hat{\rho}(k)$: Sample Autocorrelation for Lag k.

[11] Nevertheless, our main results are robust to the measure of global macro factors using Stock and Watson (2002) or the measure from the OECD database. Our method is preferred as the explanatory power of macro factors for bond yields is relatively stronger.

3.3.1 Variance Decomposition of Model Hierarchies

As mentioned in the previous section, for each country we identify two latent pricing factors, which have accounted for the majority of the variance of bond yields across all countries. It is only the global Level factor in our model that drives the national Level factors. Similarly the global Slope drives national Slope factors. Table 3.2 displays the importance of the global-level ($Share_G$) and country-specific ($Share_F$) components in Eq. (3.3) and (3.2), as well as idiosyncratic noise ($Share_X$) in Eq. (3.1), relative to the total variation in the data of seven countries. It is clear that the global factors explain the vast majority of country yields: $Share_G$ is greater than 0.6 for almost all countries.[12] Consequently, this characteristic leads us to believe the co-movement of international bond yields is generally very strong and dominates national or idiosyncratic movements.[13] The evidence is consistent with the importance of the global factors found in Diebold, Li and Yue (2008). As the global factors account for a large proportion of the information in national term structures, we are interested in the dynamics of the two global factors, Level and Slope, and seek to provide sensible economic interpretations for the factors in this study.

Although global factors clearly dominate yields, national factors remain important. The variance explained by country-specific components (i.e. $Share_F$) is non-trivial and more than two standard deviations from zero. This in turn implies, that the sum of $Share_G$ of global factors and $Share_F$ of country-specific components account for $96 - 99\%$ of bond variation across all countries.[14] The idiosyncratic noise is largely irrelevant and our model is doing a good job modeling yield co-movement. It is consistent with the early evidence of Litterman and Scheinkman (1991) for bond markets. In Appendix B.4.3, we present auxiliary analyses about global and country-level factor dynamics. Having identified significant co-movement in yields using an empirical factor approach, we now seek to model international yields in more depth in the next section.

[12] The exception is Italy potentially as those yields bear higher sovereign and hence country-specific risks.

[13] We can refer to Figure 3.1 in last section, which gives a more intuitive impression.

[14] In other words, the sum equals to the share of variance of national yield factors. Note there is a clear distinction between national factors and country-specific components. Country-specific components are the movements in national factors that are not driven by global factors.

Table 3.2: *Decomposition of Variance of Hierarchies*

Country	Posterior Mean (Std. Dev.)		
	$Share_G$	$Share_F$	$Share_X$
Italy	0.36(0.10)	0.63(0.10)	0.01(0.00)
Canada	0.71(0.07)	0.27(0.07)	0.02(0.00)
France	0.76(0.07)	0.22(0.06)	0.02(0.00)
Germany	0.74(0.07)	0.22(0.06)	0.04(0.01)
Japan	0.68(0.08)	0.30(0.07)	0.03(0.01)
UK	0.85(0.05)	0.13(0.04)	0.02(0.01)
US	0.75(0.07)	0.24(0.07)	0.01(0.00)

Notes: This table summarizes the decomposition of variance for the three-level hierarchical model of bond yields. Std. Dev. denotes the posterior standard deviation of the posterior mean. For each country, $Share_G$, $Share_F$ and $Share_X$ denote the variance shares (averaged across all maturities) of respective shocks ε_G, ε_F and ε_X at different levels. Parentheses $(\cdot)$ contain the posterior standard deviation of shares in a specific block.

3.4 Empirical Results

3.4.1 Decomposition of Structural Shocks

In Section 3.3.1, we show that the global yield factors account for the majority of the variance of bond yields. There are important co-movements of yields, although the co-movements are primarily at the long end of the yield curve according to Byrne, Fazio and Fiess (2012) and Jotikasthira, Le and Lundblad (2015).[15] Jotikasthira, Le and Lundblad (2015) suggest it is due to the uncoupling of short-term policy rates in different countries.

To evaluate the relative importance of global fundamentals and non-fundamentals in driving the co-movement of bond yields, we further decompose the 10-year Forecast Error Variance (FEV) of yields driven by innovations of global factors. Note that the shares are quantitatively similar for all countries based upon our model construction, except for Japan where inflation accounts for much less variance, i.e. around half of the shares of the other countries. Our results suggest that much of the FEV can be explained by the innovations of global factors as implied in Section 3.3.1. The remainder of FEV is explained by the country-specific components and the idiosyncratic innovations across yields at different maturities. The country-specific components in national yield factors are also meaningful as it may imply global 'spillovers', and hence we will discuss this in detail in Section 3.4.4. For now, this section focuses upon the global co-movement in yields.

In Table 3.3, we show the decomposition of the variance of all US yields explained by global factors. Our first finding is that shocks to non-fundamentals, i.e. shocks to the global Level and Slope factors, are the main sources driving interest rate movements. We observe the proportion of shocks to non-fundamentals increases with maturities. Moreover, we find that shocks to non-fundamentals are persistent and not followed by changes in fundamentals, which is consistent with Benhabib and Wang (2015). Another observation is that the Level dominates, especially at longer maturities (around 74%), whereas the Slope is relatively more important for shorter maturities (up to 18%). Diebold and Rudebusch (2013) suggest the Slope factor is connected to investors' view about the stance of current monetary policy. As suggested by Novy-Marx (2014), the underlying sources of shocks to non-fundamentals are not clear without further analysis.

Among all fundamentals, CPI accounts for a significant fraction of bond yield co-movement at shorter maturities, contributing to up to 22% of FEV of co-movement. The shares are considerably lower, however, for bonds at longer maturities. It is possible an inflation shock that imposes downward pressure on the risk premia increase short rate expectations. To have a deeper understanding of how global fundamentals affect the yields, in the next section we decompose the yield co-movements into two channels, in light of the results of Wright (2011) and Jotikasthira, Le and Lundblad (2015).

[15]Our results also support that the variance accounted for by the global yield curve increases with yield maturity, see Table 17, 18 and 19 in Appendix B.4.4.

Table 3.3: *Decomposition of US yield Variance Explained by Global Factors*

Maturity (Month)	Posterior Mean (Standard Deviation)					
	IP	CPI	PR	Level	Slope	FCI
3	0.02(0.02)	0.22(0.12)	0.03(0.03)	0.49(0.15)	0.18(0.08)	0.06(0.04)
6	0.02(0.02)	0.21(0.12)	0.03(0.03)	0.52(0.15)	0.17(0.07)	0.06(0.04)
12	0.02(0.02)	0.18(0.12)	0.03(0.02)	0.57(0.15)	0.14(0.07)	0.05(0.04)
24	0.02(0.02)	0.15(0.11)	0.02(0.02)	0.64(0.15)	0.11(0.07)	0.05(0.04)
36	0.02(0.02)	0.14(0.11)	0.02(0.02)	0.68(0.15)	0.1(0.07)	0.04(0.04)
48	0.02(0.02)	0.13(0.1)	0.02(0.02)	0.7(0.15)	0.09(0.07)	0.04(0.04)
60	0.02(0.02)	0.12(0.1)	0.02(0.02)	0.72(0.15)	0.08(0.07)	0.04(0.04)
72	0.02(0.02)	0.11(0.1)	0.02(0.02)	0.73(0.14)	0.08(0.07)	0.04(0.04)
84	0.02(0.02)	0.11(0.1)	0.02(0.02)	0.73(0.14)	0.08(0.07)	0.04(0.04)
96	0.02(0.02)	0.11(0.1)	0.02(0.02)	0.74(0.14)	0.07(0.07)	0.04(0.04)
120	0.02(0.02)	0.10(0.09)	0.03(0.02)	0.74(0.14)	0.07(0.07)	0.04(0.04)

Notes: 1. This table summarizes the posterior mean of the decomposition of 120-month Forecast Error Variance of US bond yields driven by innovations of global yield and macro factors. In each parenthesis $(\cdot)$ the posterior standard deviation of shares in a specific block is calculated from our draws, see Section 3.2. Larger Standard Deviation means higher uncertainty in the estimates, but we do not have an exact credible interval interpretation as the statistics do not necessarily follow (truncated) normal distributions.
2. IP, CPI, PR, Level, Slope and FCI denote the variance shares of shocks to global fundamentals at different maturities in the country-level block. The global fundamentals include the Industrial Production growth rate (YoY), inflation, change of policy rate (YoY), global Level, global Slope and FCI, respectively. The shares in each row sum up to 1.
3. We employ Cholesky decomposition to identify the shocks using the following ordering: IP, CPI, PR, Level, Slope and FCI. The details can be found in Appendix B.2.3.

3.4.2 Policy and Risk Compensation Channels

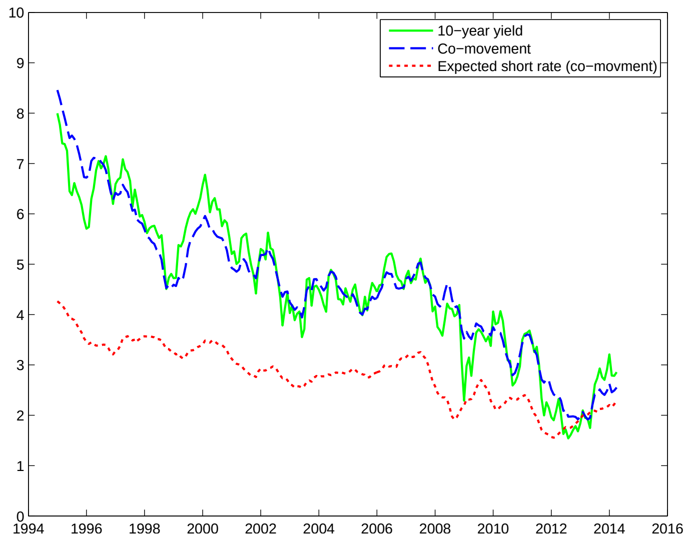
Policy rates are decided by national monetary authorities who may have inflation targets and therefore inflation explains a large proportion of short yields. Regarding the long rates, we can decompose an inflation shock into two transmission channels. The first channel is the influence on the current short rate and expected future short rates. The current short rate and future short rate expectations are closely connected to monetary policy, so we regard this channel as the 'policy channel'. The movements in this policy channel are in line with the 'Expectation Hypothesis'. The other channel is the 'risk compensation channel', through which the movements account for the bond market risk compensation for a bond at longer maturity. The compensation is also called 'term premia', which is the difference between the real long yield and the 'Expectation Hypothesis' consistent long yield. Following the approach of Wright (2011) and Jotikasthira, Le and Lundblad (2015), we aim to decompose the long yield co-movements into these two distinct channels and assess their relative importance.[16] Figure 3.2 shows the co-movement part of US 10-year long yields implied by the model and the global-driven expected short rates by decomposing the co-movement part.

In summary, the policy channel determines expected short rates while the risk compensation channel accounts for movements of the term premia. Table 3.4 displays the proportion of variance of the 10-year bond driven by global factors accounted for by each channel. The tables also show the shares of the influence of each global yield factor or global macro factor though these two alternative channels.[17] Firstly, we find that the co-movements of the 10-year bond are largely driven by the risk

[16]Our definitions of these two channels are similar to Jotikasthira, Le and Lundblad (2015), although our model structure is different. See Appendix B.2.3 for technical details.

[17]The results for other long yields (maturities 5 to 9 years) do not vary much and therefore are not displayed

Figure 3.2: *US 10-Year Bond Yields and Co-Movements*



Notes: 1. This figure shows the observed US 10-year bond yields, and the global-driven yield movements and term premia implied by the model. The 10-year nominal yields are plotted by the solid line. The dashed line plots the portion of yields driven by global factors (co-movement), and the dotted line is the term premia part in the co-movement. The unit is percentage.

compensation channel. For all seven countries, this risk channel accounts for more than 53% of the total variance of long rate co-movement. For Japan, the risk compensation channel even accounts for 96%. The relative importance of the risk compensation channel is in line with the results in Jotikasthira, Le and Lundblad (2015).[18] Secondly, we find that inflation is very important in driving the global co-movement of long yields through both channels. Take the US for example, recall Table 3.3, the joint contribution of CPI inflation to the 10-year bond co-movement is only 10%. But when we decompose the influence into the policy channel and risk compensation channel, the contribution through each channel is significantly increased, especially for the policy channel through which the share triples.

Why might this be the case? Impulse responses help us understand why. There is a sizable reduction in the overall influence of inflation because risk and policy channels counteract one another, which also applies to other global fundamentals. We plot the impulse response of the co-movement of the US 10-year bond to global shocks in Figure 3.3 and the two offsetting effects are revealed. A global macro shock that has positive effects through the policy channel usually has negative effects through the risk compensation channel. These opposite effects of macroeconomic shocks (i.e. IP, CPI, PR and FCI) can potentially explain why the changes in yield factors are not primarily driven by the macroeconomic shocks. This observation highlights the importance of shocks to non-fundamentals, and we need to go beyond a fundamentals-driven model to capture bond yield movements.

here.

[18] Jotikasthira, Le and Lundblad (2015) indicate the risk compensation channel accounts for around 80% and 42% for the US and Germany, respectively. We include the financial crisis period in our sample so we have a decreased share for the US and an increased share for Germany.

Table 3.4: *Decomposition of Variance through Two Channels (10-Year Bonds)*

Country	Channel	Posterior Mean (Std. Dev.)					
		IP	CPI	PR	Level	Slope	FCI
US	Policy	0.02	0.30	0.04	0.48	0.09	0.07
	47%	(0.02)	(0.19)	(0.04)	(0.24)	(0.1)	(0.07)
	Risk Compensation	0.02	0.14	0.05	0.66	0.07	0.06
	53%	(0.02)	(0.11)	(0.03)	(0.17)	(0.06)	(0.05)
UK	Policy	0.02	0.32	0.04	0.46	0.09	0.07
	39%	(0.02)	(0.19)	(0.04)	(0.24)	(0.1)	(0.07)
	Risk Compensation	0.02	0.11	0.05	0.71	0.06	0.05
	61%	(0.02)	(0.09)	(0.03)	(0.15)	(0.06)	(0.05)
Germany	Policy	0.02	0.33	0.04	0.45	0.09	0.07
	23%	(0.02)	(0.19)	(0.04)	(0.24)	(0.1)	(0.07)
	Risk Compensation	0.02	0.08	0.04	0.77	0.06	0.04
	77%	(0.02)	(0.06)	(0.02)	(0.11)	(0.06)	(0.04)
France	Policy	0.02	0.32	0.04	0.46	0.09	0.07
	33%	(0.02)	(0.19)	(0.04)	(0.24)	(0.1)	(0.07)
	Risk Compensation	0.02	0.10	0.05	0.74	0.06	0.05
	67%	(0.02)	(0.08)	(0.02)	(0.13)	(0.06)	(0.04)
Italy	Policy	0.02	0.30	0.04	0.48	0.09	0.07
	29%	(0.02)	(0.19)	(0.04)	(0.24)	(0.1)	(0.07)
	Risk Compensation	0.02	0.08	0.04	0.76	0.06	0.04
	71%	(0.02)	(0.07)	(0.02)	(0.12)	(0.06)	(0.04)
Canada	Policy	0.02	0.30	0.04	0.48	0.09	0.07
	27%	(0.02)	(0.19)	(0.04)	(0.24)	(0.1)	(0.07)
	Risk Compensation	0.02	0.08	0.04	0.76	0.06	0.04
	73%	(0.02)	(0.07)	(0.02)	(0.12)	(0.06)	(0.04)
Japan	Policy	0.02	0.22	0.03	0.58	0.09	0.06
	4%	(0.02)	(0.17)	(0.04)	(0.23)	(0.1)	(0.07)
	Risk Compensation	0.02	0.06	0.03	0.81	0.06	0.03
	96%	(0.02)	(0.06)	(0.01)	(0.09)	(0.06)	(0.03)

Notes: 1. This table summarizes the decomposition of 120-month Forecast Error Variance of the 10-year bond yields driven by innovations of factors through two channels: the policy and risk premia channels. In each parenthesis $(\cdot)$ the posterior standard deviation of shares in a specific block is calculated from our draws, see Section 3.2. Larger standard deviation means higher uncertainty in the estimates, but we do not have an exact credible interval interpretation as the statistics do not necessarily follow (truncated) normal distributions.
2. IP, CPI, PR, Level, Slope and FCI denote the variance shares of shocks to global fundamentals at different maturities in the country-level block. The global fundamentals include the Industrial Production growth rate (YoY), inflation, change of policy rate (YoY), global Level, global Slope and FCI, respectively. The shares in each row sum up to 1.
3. We employ Cholesky decomposition to identify the shocks using the following ordering: IP, CPI, PR, Level, Slope and FCI. The details can be found in Appendix B.2.3.

Wright (2011) suggests that inflation uncertainly is closely connected to global bond yield movements, through both the policy and the risk compensation channels. Moreover, Wright (2011) points out that the term premia is positively correlated with inflation expectation, and our findings confirm this mechanism. The top-right panel of Figure 3.3 indicates that a positive shock on current inflation immediately drives up the term premia, as the increase in inflation may raise inflation uncertainty and hence the risk premia. Inflation is the most important driver of the term premia among all macro variables, both in terms of the quantity and persistency.

Why might the term premium be sensitive to inflation? It is worth mentioning a positive financial shock drives down the term premia, and the effects are quite persistent; see the bottom-right panel of Figure 3.3. The underlying mechanism is that a positive or malignant financial shock will drive down future long-run inflation, and hence this expectation lowers the term premia.[19]

Figure 3.3 also shows how the global policy channel reacts to changes in inflation. Inflation shocks are very persistent when compared with other macro variables. A positive shock to global inflation is accompanied by a decrease in the policy rate, which seems to contrast a standard Taylor rule. This observation is because for the G7 countries, our identified global inflation shock is not orthogonal to changes in global bond yield factors, and a positive inflation shock is in fact a negative shock to global Level factor. A potential explanation is to interpret this shock as a cost-push or markup shock. When the positive inflation shock is accompanied by undesired movements in Industrial Production (IP) growth rate or FCI, the global policy rate level may decrease to offset to these expected movements.

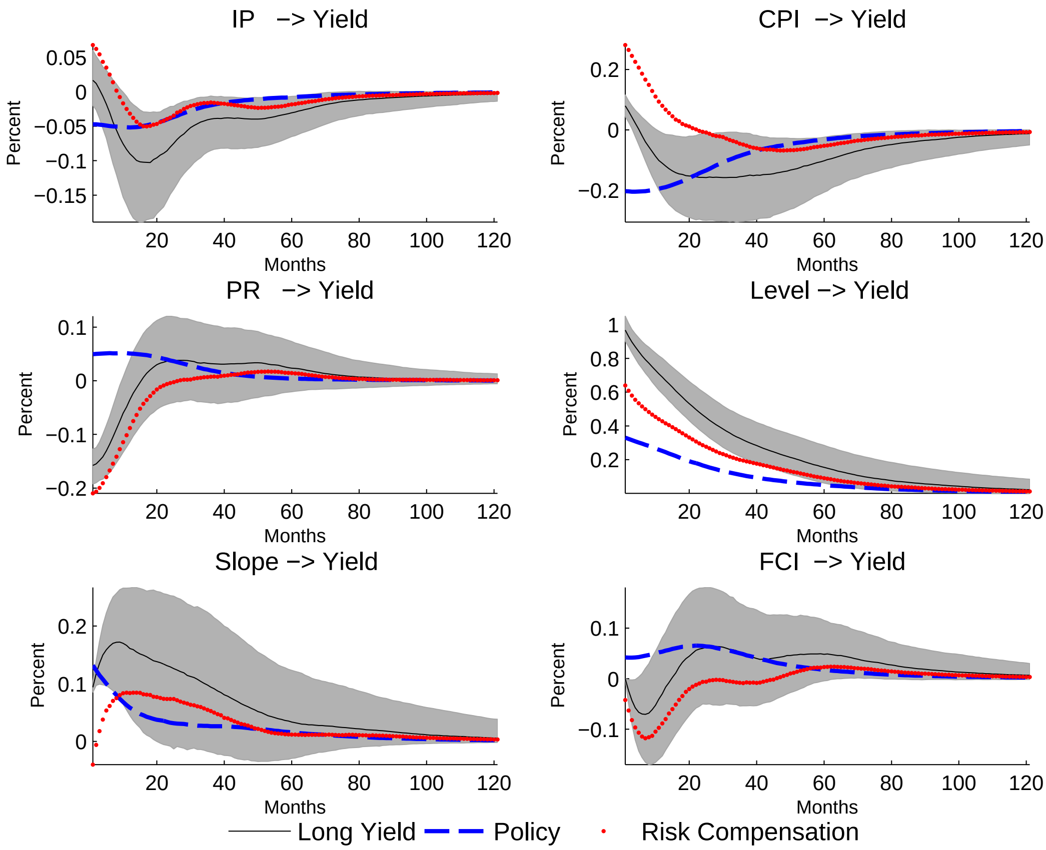
It is worth emphasizing our results are obtained from the identification scheme without imposing any hard restrictions such that global macro factors are forced to be pricing factors. In fact, we will show later the results are robust if the global macro variables are treated as pricing factors, i.e. spanned macro factors. As we have discussed, this is because the method of Moench, Ng and Potter (2013) identifies the factors by allowing for the time-series properties of the underlying global dynamics, so the information of macro variables is naturally incorporated without imposing further restrictions.

3.4.3 What Drives Non-Fundamental Co-Movements?

We have identified global yield factors using a parsimonious specification, and the factor dynamics are considered more plausible with the augmentation of fundamentals. An interesting follow-up question is what the non-fundamentals are, as they contribute a large proportion of bond yield co-movements. We appeal to two possible explanations about non-fundamental movements that are well documented in the literature. The first explanation corresponds to the sentiment-based theory favored by Kumar and Lee (2006), Bansal and Shaliastovich (2010) and Benhabib and Wang (2015). As suggested in Ludvigson (2004), consumer confidence index is a widely used measure of investor sentiment. We

[19] We find that inflation news shocks are closely related to the FCI current shocks, and a shock to inflation will induce unfavorable movements in Industrial Production (IP) growth rate or FCI. The results are not reported here for the sake of brevity but are available upon request.

Figure 3.3: *Impulse Responses of US 10-Year Bond Co-Movement to Global Shocks*



Notes: 1. This figure decomposes structural shocks to global factors that cause one percentage point increase in the US 10-year bond. The solid lines in the above panels show the impulse responses of 10-year long yield movements to six orthogonal (positive) global shocks. Cholesky decomposition is employed to identify the shocks. The 16 to 84 percent posterior coverage intervals for the long yield are are shaded in gray.
2. Each shock can be further decomposed into two channels: the policy channel (blue dashed line) and the risk compensation channel (red dotted line). See Appendix B.2.3 for technical details.

obtain leading indicator aggregates of G7 from the OECD database as proxies of global sentiment, which include the composite leading indicator, business confidence index and consumer confidence index. Alternatively, asset prices can be driven by economic uncertainty, see Bloom (2014) for a comprehensive review. We use the US and Europe economic policy uncertainty indicators constructed by Baker, Bloom and Davis (2013) as the measure of economic uncertainty.

Table 3.5 reports the regression results about global co-movements. The regression of global Level factor on global macro factors used in this paper shows only a relatively smaller portion of variance is driven by macro factors (around 20%), which is consistent with our previous findings. Adding sentiment and/or economic uncertainty measures greatly increases the explanatory power, and the adjusted R^2 is increased by more than 50%. With respect to the global Slope factor, it seems only the sentiment measures can significantly increase the adjusted R^2, as the global Slope has been explained by macroeconomic information to a large degree.

Table 3.5: *Co-Movement Regressions*

	CLI^{G7}	BCI^{G7}	CCI^{G7}	PU^{US}	PU^{EU}	M+Constant	$adjR^2$
Level						*	20.14%
	-0.35(0.08)	-0.07(0.09)	0.89(0.05)			*	65.22%
				-0.01(0.00)	-0.01(0.00)	*	58.74%
	-0.48(-0.01)	-0.07(-0.01)	0.47(0.08)	0.08(0.00)	0.06(0.00)	*	72.96%
Slope						*	42.28%
	0.60(0.05)	-0.42(0.06)	-0.03(0.03)			*	60.72%
				0.00(0.00)	0.00(0.00)	*	42.72%
	0.62(0.00)	-0.43(0.00)	0.02(0.05)	0.06(0.00)	0.05(0.00)	*	60.80%
TP	-0.17(0.00)	-0.12(0.00)	0.27(0.04)	0.04(0.00)	0.03(0.00)	*	93.76%
y^E	-0.41(-0.01)	-0.04(0.00)	0.36(0.06)	0.07(0.00)	0.05(0.00)	*	83.47%

Notes: This table summarizes the regressions of global Level and Slope factors, and the US 10-year term premia (TP) and long-term short rate expectations y^E, on global macro variables, leading indicators and/or policy uncertainty indicators. M collects global macro variables used in our models. The leading indicators are G7 aggregates from the OECD database, where *CLI*, *BCI* and *CCI* are the composite leading indicator, business confidence index and consumer confidence index, respectively. Policy uncertainty indicators include the US policy uncertainty index PU^{US} and the Europe policy uncertainty index PU^{EU}, which are calculated by Baker, Bloom and Davis (2013). The sample is from 1994:12 to 2014:03 at monthly frequency. The standard errors are given in parentheses $(\cdot)$ and the Adjusted R_2 are reported.

In Table 3.5 we also report the regressions of the global-driven movements of the US 10-year bond through two channels. The results for other countries or at other maturities are very similar, as the global-driven movements of all countries are linear functions of global factors. All measures of sentiment and economic uncertainty are highly significant. With the strikingly high explanatory power for the movements through both channels, we are reassured that non-fundamental movements are indeed closely related to investor sentiment and economic uncertainty. This important finding parallels the fast-growing literature with the consideration of sentiment or economic uncertainty. The co-movement in global yield curves can be almost exclusively characterized by the information of macro variables, sentiment and economic uncertainty, and therefore a structural model with these salient features can offer a satisfactory explanation for the changes in global bond markets.

3.4.4 Contagion

Apart from the global shocks, how would the country-specific components of one country affect other countries? In our model, 'spillover effects' contagion arises after controlling for shocks driven by the common factors. As we have discussed, common shocks are significantly related to changes in sentiment and economic uncertainty, which are well captured by our identified global Level and Slope factors augmented with global macro information. Given the common shocks have been controlled for, we are able to analyze the 'spillovers', which are induced by country-specific components and may be asymmetric among our sample of countries. Jotikasthira, Le and Lundblad (2015) suggest country-specific components are caused by the uncoupling of policy rates, so this analysis can help us understand the spillovers of diverging monetary policies, which may be closely related to country-specific fundamentals.[20]

Our constructed model allows us to separate the country-specific components driving the national yield factors from the global yield factors. Firstly, we would like reiterate the covariance structure in our model. For each country-specific component, its innovations are assumed serially uncorrelated. However, our model is silent on the cross-equation correlations *between* different components. That is, the equations of the dynamics of country-specific components are *seemingly unrelated*, so potential correlations *between* country-specific components are not precluded. In fact, if these identified components are truly correlated, the interdependent relations imply 'spillovers' among countries apart from the common shocks.

It is evident that there exist strong cross country correlations.[21] For example, US Level and Slope factors are related to Canadian, German, Italian, and Japanese factors. German factors are also related France and Italy. The strong correlations encourage us to explore the inner mechanism of potential 'spillovers'. How would the country-specific components in yield factors of one country affect the movements of the components of another country and to what extent? Are these effects symmetric or asymmetric? The answers for these questions are desirable and we will conduct the following evaluation process and try to provide sensible evidence.

We would like to further analyze the global connectedness by quantifying the 'spillover effects' among countries, so we employ the approach proposed by Diebold and Yilmaz (2009, 2014); see Appendix B.2.4 for details. In other words, we construct a VAR(1) system using the country-specific components, and then conduct generalized variance decomposition of the form proposed

[20]Our empirical results affirm that the country-specific components influencing the national yield factors are largely accounted by the divergence of policy rates in different countries. We can construct the indicators of the divergence of policy rates by subtracting the principal component of all policy rate series from each national policy rate series, and the residuals indicate the monetary policy divergence. For each country, adding the indicator of divergence as an additional explainable variable in the regressions of global yield factors can greatly improve the explanatory power of the regression model, and the usefulness of this local divergence variable is distinguished by the high significance. The finding is robust as it holds for all countries across yield maturities, especially for the short yields. The results are are available upon request.

[21]Table 20 in Appendix B.5 displays the correlation matrix of the country-specific components in national Level and Slope factors.

by Koop, Pesaran and Potter (1996). The decomposition helps us delineate connectedness, because this arises through the covariance matrix that can reveal contemporaneous correlation.[22] To quantify connectedness we follow Diebold and Yilmaz (2009, 2014) and calculate Spillover Indexes based on the variance decomposition. The details are reported in Table 3.6.

According to the spillover table, we find that the movements in bond markets of France, Germany, UK and US are susceptible to changes in other countries. Moreover, the bond markets of Italy, Canada, France and Germany contribute to large proportions of changes in other countries. We find that around one fourth of the variance across home countries is due to the shocks on the country-specific components of foreign countries.

To have a more intuitive understanding of the asymmetric spillovers among the country-specific components, we use the evidence in Table 3.6 to plot the network graphs in Figure 3.4. This graph displays idiosyncratic connections based upon the distance and the thickness of connections.[23] We find that there are two main clusters: Europe and North America. The European markets are united, but the UK is relatively unconnected to Europe. In general the UK market has a similar link to European markets and the markets of North America. The Japanese market is also very independent, as the edges connected to the node of Japan are relatively thin and hence the node is further away from the other clusters. We can see there is a large distance between the US and Italy, and it seems there is no significant direct connection between these two countries. This means the country-specific components of Italy are not likely to directly affect the components of the US, implying there might not be strong spillover effects of the sovereign crisis from Italy to the US through a direct channel. However, if the sovereign crisis affects all European markets, the US market will also be influenced, but the spillovers may boil down to global co-movements that we have discussed in the last section.

By construction, the directions and quantity of the 'spillover effects' contagion are identified, which can be used for further analysis of the network model. The measured asymmetries in contagion effects controlling for co-movement are by far new to related research of financial contagion.

[22]Alternative schemes, for example, network connectedness measures based on Granger-causal patterns of Billio et al. (2012), can be employed as robustness checks. See Appendix B.5 for details.

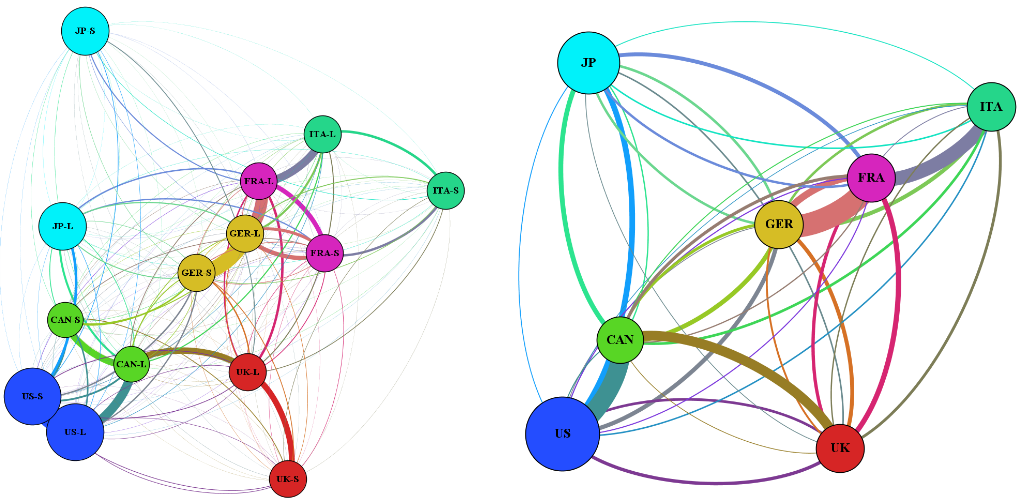
[23]The node size is determined by the amount of total debt securities outstanding as of June 2014. The node location is determined by the *ForceAtlas2* algorithm of Jacomy et al. (2014): We assume that nodes repel each other, but edges attract the nodes according to average pairwise directional connectedness 'to' and 'from' in Table 3.6. The algorithm finds a steady state in which repelling and attracting forces reach a balance.

Table 3.6: *Spillover Table of the Country-Specific Components*

To	From															
	ITA_L	ITA_S	CAN_L	CAN_S	FRA_L	FRA_S	GER_L	GER_S	JP_L	JP_S	UK_L	UK_S	US_L	US_S	Contribution From Others	
ITA_L	96.31	0.00	0.26	0.54	0.02	0.44	0.89	0.82	0.06	0.11	0.24	0.18	0.06	0.08	3.69	9.52
ITA_S	7.08	87.09	0.00	0.32	0.21	0.14	2.46	0.20	0.58	0.09	0.36	1.17	0.00	0.30	12.91	
CAN_L	3.14	0.48	90.17	0.28	0.98	0.17	0.47	1.40	1.79	0.05	0.53	0.02	0.27	0.24	9.83	18.76
CAN_S	0.95	0.10	19.22	71.57	0.08	0.45	2.70	4.17	0.00	0.01	0.00	0.02	0.71	0.03	28.43	
FRA_L	22.37	0.00	2.12	0.78	61.73	0.25	2.17	1.57	3.89	0.00	4.04	0.09	0.02	0.98	38.27	64.04
FRA_S	0.01	6.61	0.93	1.91	12.98	60.99	8.42	2.06	3.52	0.08	1.29	0.75	0.43	0.01	39.01	
GER_L	5.74	0.20	2.04	0.64	34.03	12.14	37.55	1.66	3.21	0.01	2.09	0.00	0.02	0.67	62.45	69.27
GER_S	0.08	1.63	2.42	0.22	0.16	1.57	43.48	48.04	1.28	0.06	0.30	0.67	0.04	0.05	51.96	
JP_L	0.31	0.65	5.76	2.17	0.12	0.01	0.02	2.63	87.05	0.13	0.03	0.53	0.58	0.01	12.95	21.86
JP_S	0.14	1.21	0.87	0.62	3.08	0.97	0.58	1.11	0.39	90.57	0.07	0.03	0.03	0.33	9.43	
UK_L	2.67	2.01	18.17	0.00	6.90	2.07	0.23	3.79	0.01	2.25	58.47	0.02	1.28	2.14	41.53	50.12
UK_S	0.02	0.38	0.04	2.09	0.00	1.22	1.26	2.20	0.12	0.02	16.47	74.92	1.10	0.17	25.08	
US_L	1.24	0.00	25.34	4.21	0.24	0.52	0.00	3.61	1.25	0.97	1.12	1.54	59.96	0.02	40.04	68.92
US_S	0.96	0.04	5.17	4.95	0.31	0.27	0.54	4.00	7.44	1.60	1.92	1.70	39.74	31.36	68.64	
Contribution	44.70	13.32	82.37	18.72	59.12	20.22	63.22	29.22	23.54	5.36	28.46	6.70	44.28	5.02	**444.22**	
To others	50.93		81.58		66.10		47.29		28.39		18.66		9.54			**302.48**
Including own	141.01	100.41	172.53	90.29	120.85	81.21	100.77	77.25	110.59	95.94	86.93	81.62	104.23	36.39	Spillover Index	
															$= 31.7\%$	$= 24\%$

Notes: 1. This table summarizes the spillover table of the country-specific components among the Level and Slope factors of all countries: Italy (ITA), Canada (CAN), France (FRA), Germany (GER), Japan (JP), the UK and the US. Subscripts *L* and *S* are for Level and Slope factors respectively.
2. The underlying variance decomposition (reported in percentage) is based upon a monthly VAR of order 1, identified using a generalized variance decomposition. The (i, j)-th value is the estimated contribution to the variance of the 12-month-ahead forecast error of country-specific component i coming from innovations to the component j.
3. The spillover index is the cross variance share, i.e. the variance due to the shocks on $j, j \neq i$ relative to total forecast error variation of i. Two indexes that measure spillovers among components and countries are calculated, respectively. See Appendix B.2.4 for details.
4. The final two columns set out the respective fractions of movements in a component caused by shocks to other components and of that in a country caused by shocks to other countries. The final three rows respectively set out the contribution of a component to other components, a country to other countries or the total contribution to components including to its own.

Figure 3.4: *Network of Global Spillovers*



Notes: The left panel shows the spillovers of country-specific components among seven countries, which is constructed according to the results in Table 3.6. In the right panel, the spillover effects of the same country are grouped. The mnemonics are defined as in Table 3.6.

3.5 Robustness

3.5.1 Macro Spanning Restrictions

As we have discussed, one key contribution to the current literature is that we introduce a flexible identification scheme robust to model specification, in particular, the macro spanning restrictions. Although it seems we employ a seemingly unspanned setup, our results are in fact not sensitive to these restrictions. If macro information is truly spanned by bond yields, then our identified factors are naturally close to rotations of macro factors and hence can satisfactorily span the macros.

To validate the above argument, we proceed with a robustness check by allowing global macro factors to be pricing factors. Equation (3.1) now becomes

$$X_{ibt} = \Lambda_{ib}^{F} F_{bt} + \Lambda_{ib}^{M} M_t + e_{ibt}^{X}.$$

We are curious about to what extent the macro spanning condition affects bond yields, as macro factors now have direct influence. Figure 3.5 sets out the impulse responses and Table 3.7 provides a quantitative evaluation of structural shocks in a spanned setup. Not surprisingly, spanned and unspanned setups give qualitatively indistinguishable and quantitatively similar results because of the reasons we discussed.

Table 3.7: *Decomposition of Variance through Two Channels (US 10-Year Bonds)*

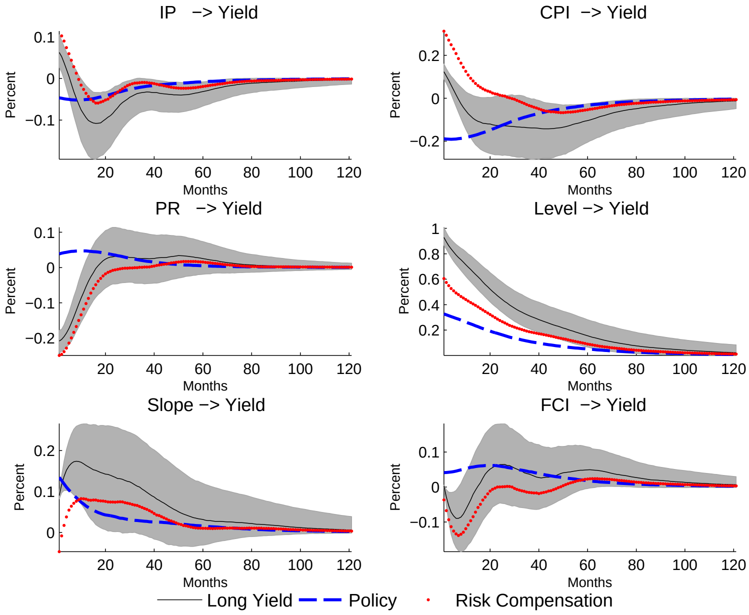
Country	Channel	Posterior Mean (Std. Dev.)					
		IP	CPI	PR	Level	Slope	FCI
US	Policy	0.02	0.28	0.04	0.50	0.09	0.07
	46%	(0.02)	(0.19)	(0.04)	(0.24)	(0.10)	(0.07)
	Risk Compensation	0.02	0.14	0.05	0.66	0.07	0.06
	54%	(0.02)	(0.11)	(0.03)	(0.17)	(0.06)	(0.05)

Notes: 1. This table summarizes the decomposition of 120-month Forecast Error Variance of the US 10-year bond yields driven by innovations of factors through two channels: the policy and risk premia channels. In each parenthesis $(\cdot)$ the posterior standard deviation of shares in a specific block is calculated from our draws, and the macro spanning condition is imposed. We employ Cholesky decomposition to identify the shocks using the following ordering: IP, CPI, PR, Level, Slope and FCI. The details can be found in Appendix B.2.3.

3.5.2 Results Excluding the Zero Lower Bound Episode

During the financial crisis, the nominal short-term interest rates of the US and the EU drop to near zero. The zero lower bound limits the capacity of the central banks to react to economic conditions, and possibly weakens the explanatory power of macroeconomic fundamentals. Therefore, we use a subsample of around 15 years $(1994 : 12 - 2008 : 04)$ to conduct a robustness check. Table 3.8 reports the variance decomposition of model hierarchies. One finding worth mentioning is that the

Figure 3.5: *Impulse Responses of US 10-Year Bond Yields to Global Shocks (Macro Spanning)*



Notes: 1. This figure decomposes structural shocks to global factors that cause one percentage point increase in the US 10-year bond with the macro spanning condition. The solid lines in the above panels show the impulse responses of 10-year long yield movements to six orthogonal (positive) global shocks. Cholesky decomposition is employed to identify the shocks. The 16 to 84 percent posterior coverage intervals for the long yield are are shaded in gray.

2. Each shock can be further decomposed into two channels: the policy channel (blue dashed line) and the risk compensation channel (red dotted line). See Appendix B.2.3 for technical details.

bond yields of Italy are almost exclusively driven by global factors. The global-driven share of the US is relatively smaller than the full sample estimate, which implies that US plays an important role in driving the co-movement during the financial crisis.

Table 3.8: *Decomposition of Variance of Hierarchies (Subsample)*

Country	Posterior Mean (Std. Dev.)		
	$Share_G$	$Share_F$	$Share_X$
Italy	0.91(0.09)	0.08(0.09)	0.00(0.00)
Canada	0.75(0.17)	0.23(0.16)	0.01(0.01)
France	0.84(0.14)	0.14(0.13)	0.01(0.01)
Germany	0.75(0.17)	0.23(0.16)	0.02(0.01)
Japan	0.64(0.21)	0.35(0.21)	0.01(0.00)
UK	0.76(0.16)	0.22(0.15)	0.02(0.01)
US	0.57(0.22)	0.42(0.22)	0.01(0.00)

Notes: This table summarizes the decomposition of variance for the three-level hierarchical model of bond yields. Std. Dev. denotes the posterior standard deviation of the posterior mean. For each country, $Share_G$, $Share_F$ and $Share_X$ denote the variance shares (averaged across all maturities) of respective shocks ε_G, ε_F and ε_X at different levels. Parentheses $(\cdot)$ contain the posterior standard deviation of shares in a specific block. The sample period is $1994:12-2008:04$.

Table 3.9 shows macro fundamentals have much stronger explanatory power before the financial crisis, especially for short yields. It shows macro fundamentals are important ingredients in determining policy rates in normal times. The results are obtained with unspanned restrictions, which reiterate the robustness of our identification scheme. Non-fundamental shocks, through weakened, still account for more than half of the variance of long yields. When compared with the full sample results, these results suggest sentiment and economic uncertainty play essential roles during the financial crisis. These empirical results leave a question open: Whether global crisis transmissions can be explained in a self-fulfilling mechanism? We expect further research in this direction.

Lastly, we notice the subsample results are subject to the small sample problem indicated by Bauer and Hamilton (2015). The posterior coverage intervals for the impulse responses to global shocks can be implausibly wide, which implies the existence of observationally equivalent global dynamics. This evidence parallels the findings of Bauer and Hamilton (2015) that small-sample macroeconomic effects on bond yields may not be robust. The usage of tight priors with economic restrictions can potentially reconcile the model with this evidence, but it is beyond the scope of this paper.

Table 3.9: *Decomposition of US yield Variance Explained by Global Factors (Subsample)*

Maturity (Month)	Posterior Mean (Standard Deviation)					
	IP	CPI	PR	Level	Slope	FCI
3	0.16(0.09)	0.12(0.09)	0.07(0.05)	0.19(0.13)	0.23(0.12)	0.22(0.12)
6	0.16(0.09)	0.12(0.09)	0.07(0.05)	0.21(0.13)	0.23(0.12)	0.21(0.12)
12	0.15(0.09)	0.12(0.1)	0.07(0.05)	0.23(0.14)	0.22(0.12)	0.2(0.11)
24	0.15(0.09)	0.13(0.1)	0.07(0.05)	0.28(0.14)	0.2(0.12)	0.17(0.11)
36	0.14(0.09)	0.13(0.1)	0.07(0.05)	0.32(0.15)	0.19(0.11)	0.15(0.1)
48	0.14(0.09)	0.13(0.11)	0.07(0.05)	0.34(0.16)	0.18(0.11)	0.13(0.1)
60	0.14(0.09)	0.14(0.11)	0.07(0.05)	0.36(0.17)	0.17(0.11)	0.12(0.09)
72	0.14(0.09)	0.14(0.11)	0.07(0.05)	0.36(0.18)	0.17(0.11)	0.12(0.09)
84	0.14(0.1)	0.14(0.11)	0.07(0.05)	0.37(0.18)	0.17(0.11)	0.12(0.09)
96	0.14(0.1)	0.14(0.11)	0.07(0.05)	0.37(0.18)	0.16(0.11)	0.11(0.09)
120	0.14(0.1)	0.14(0.11)	0.07(0.05)	0.38(0.19)	0.16(0.11)	0.11(0.09)
Channel	10-Year Bond					
Policy	0.18	0.20	0.07	0.23	0.17	0.15
46%	(0.12)	(0.15)	(0.07)	(0.17)	(0.14)	(0.14)
Risk Compensation	0.16	0.14	0.08	0.28	0.21	0.13
54%	(0.1)	(0.11)	(0.06)	(0.17)	(0.14)	(0.1)

Notes: 1. This table summarizes the posterior mean of the decomposition of 120-month Forecast Error Variance of US bond yields driven by innovations of global yield and macro factors. The lower subtable displays decomposition of 10-year bond into two channels. In each parenthesis $(\cdot)$ the posterior standard deviation of shares in a specific block is calculated from our draws, see Section 3.2. The sample period is $1994:12-2008:04$. We employ Cholesky decomposition to identify the shocks using the following ordering: IP, CPI, PR, Level, Slope and FCI. The details can be found in Appendix B.2.3.

3.6 Conclusion

We propose a new ‘Fundamentals-Augmented Hierarchical Factor Model’ to jointly identify global and national Level and Slope factors augmented with global fundamentals: inflation, real activity, changes in policy rate and financial conditions. Co-movement accounts for on average two thirds of variability in global bond yields. Our method is robust to the macro spanning condition and able to recover significant explanatory power of global inflation shocks for global yield co-movement, through a policy channel and a risk compensation channel. Shocks to non-fundamentals are persistent and account for the majority of global term structure movement. Moreover, we find that the non-fundamental movements can be satisfactorily explained by measures of sentiment and economic uncertainty. Country-specific components contribute to the majority of remaining variance, of which one fourth is due to spillovers.

There are many possible avenues for future work. As the information driving bond yields is properly labeled in this paper, it is especially desirable to propose a structural model with the consideration of sentiment and economic uncertainty to explain global transmissions. We notice ‘spillover effects’ are mainly caused by divergence in policy rates, but it may also be interesting to specifically evaluate whether the contagion across different countries is related to fundamental or nonfundamental drivers. This paper does not explicitly model potential time-varying nonlinear dynamics of yield factors such as regime shifts. Allowing for nonlinearity can be promising in unfolding more informative dynamics of fundamental and non-fundamental fluctuations.

Bibliography

Abbritti, Mirko, Salvatore Dell'Erba, Antonio Moreno, and Sergio Sola. 2013. "Global factors in the term structure of interest rates." International Monetary Fund International Monetary Fund Working Paper WP/13/223.

Adrian, Tobias, Richard K. Crump, and Emanuel Moench. 2013. "Pricing the term structure with linear regressions." *Journal of Financial Economics*, 110(1): 110–138.

Altavilla, Carlo, Raffaella Giacomini, and Giuseppe Ragusa. 2014. "Anchoring the Yield Curve Using Survey Expectations." European Central Bank Working Paper Series 1632.

Andrews, Donald W.K., and Werner Ploberger. 1994. "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica*, 62(6): 1383–1414.

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

Anh, Le, and Scott Joslin. 2013. "Interest rate volatility and no-arbitrage affine term structure models." University of North Carolina at Chapel Hill Working Paper.

Avramov, Doron. 2002. "Stock return predictability and model uncertainty." *Journal of Financial Economics*, 64(3): 423–458.

Avramov, Doron, and Tarun Chordia. 2006. "Asset pricing models and financial market anomalies." *Review of Financial Studies*, 19(3): 1001–1040.

Bacchetta, Philippe, and Eric Van Wincoop. 2004. "A scapegoat model of exchange-rate fluctuations." *American Economic Review*, 94(2): 114–118.

Bai, Jushan, and Peng Wang. 2015. "Identification and Bayesian estimation of dynamic factor models." *Journal of Business and Economic Statistics*, 33(2): 221–240.

Bai, Jushan, and Serena Ng. 2006. "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions." *Econometrica*, 74(4): 1133–1150.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2013. "Measuring economic policy uncertainty." The University of Chicago Booth School of Business Working Paper.

Bansal, Naresh, Robert A. Connolly, and Chris Stivers. 2014. "The stock-bond return relation, the term-structure's slope, and asset-class risk dynamics." *Journal of Financial and Quantitative Analysis*, FirstView: 1–44.

Bansal, Ravi, and Ivan Shaliastovich. 2010. "Confidence risk and asset prices." *American Economic Review*, 100(2): 537–41.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler. 2005. "Comovement." *Journal of Financial Economics*, 75(2): 283–317.

Bauer, Gregory H., and Antonio Diez de los Rios. 2012. “An international dynamic term structure model with economic restrictions and unspanned risks.” Bank of Canada Working Paper.

Bauer, Michael D. 2015. “Restrictions on risk prices in dynamic term structure models.” Federal Reserve Bank of San Francisco Working Paper.

Bauer, Michael D., and Glenn D. Rudebusch. 2015. “Resolving the spanning puzzle in macro-finance term structure models.” Federal Reserve Bank of San Francisco Working Paper.

Bauer, Michael D., and James D. Hamilton. 2015. “Robust bond risk premia.” University of California at San Diego Working Paper.

Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu. 2012. “Correcting estimation bias in dynamic term structure models.” *Journal of Business and Economic Statistics*, 30(3): 454–467.

Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu. 2014. “Comment on ‘term premia and inflation uncertainty: empirical evidence from an international panel dataset’.” *American Economic Review*, 104(1): 323–337.

Benhabib, Jess, and Pengfei Wang. 2015. “Private information and sunspots in sequential asset markets.” *Journal of Economic Theory*, 158, Part B: 558 – 584.

Benveniste, Albert, Michel Métivier, and Pierre Priouret. 1990. *Adaptive Algorithms and Stochastic Approximations.* Springer.

Bernanke, Ben, Mark Gertler, and Simon Gilchrist. 1996. “The financial accelerator and the flight to quality.” *The Review of Economics and Statistics*, 78(1): 1–15.

Bernanke, Ben S., and Jean Boivin. 2003. “Monetary policy in a data-rich environment.” *Journal of Monetary Economics*, 50(3): 525–546.

Bianchi, Francesco, Haroon Mumtaz, and Paolo Surico. 2009. “The great moderation of the term structure of UK interest rates.” *Journal of Monetary Economics*, 56(6): 856–871.

Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon. 2012. “Econometric measures of connectedness and systemic risk in the finance and insurance sectors.” *Journal of Financial Economics*, 104(3): 535–559.

Björk, Tomas, and Bent Jesper Christensen. 1999. “Interest rate dynamics and consistent forward rate curves.” *Mathematical Finance*, 9(4): 323–348.

Bloom, Nicholas. 2014. “Fluctuations in uncertainty.” *The Journal of Economic Perspectives*, 28(2): 153–175.

Brandt, Michael W., Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud. 2005. “A simulation approach to dynamic portfolio choice with an application to learning about return predictability.” *Review of Financial Studies*, 18(3): 831–873.

Byrne, Joseph P., Giorgio Fazio, and Norbert Fiess. 2012. “Interest rate co-movements, global factors and the long end of the term spread.” *Journal of Banking and Finance*, 36(1): 183–192.

Cagetti, Marco, Lars Peter Hansen, Thomas Sargent, and Noah Williams. 2002. “Robustness and pricing with uncertain growth.” *Review of Financial Studies*, 15(2): 363–404.

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

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

Carriero, Andrea, and Raffaella Giacomini. 2011. "How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?" *Journal of Econometrics*, 164(1): 21–34.

Carriero, Andrea, George Kapetanios, and Massimiliano Marcellino. 2012. "Forecasting government bond yields with large Bayesian vector autoregressions." *Journal of Banking & Finance*, 36(7): 2026–2047.

Carter, Chris K., and Robert Kohn. 1994. "On Gibbs sampling for state space models." *Biometrika*, 81(3): 541–553.

Chen, Zengjing, and Larry Epstein. 2002. "Ambiguity, risk, and asset returns in continuous time." *Econometrica*, 70(4): 1403–1443.

Christensen, Jens H.E., and Glenn D. Rudebusch. 2012. "The response of interest rates to US and UK quantitative easing." *The Economic Journal*, 122(564): F385–F414.

Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch. 2011. "The affine arbitrage-free class of Nelson-Siegel term structure models." *Journal of Econometrics*, 164(1): 4–20.

Christensen, Jens H.E., Jose A. Lopez, and Glenn D. Rudebusch. 2010. "Inflation expectations and risk premiums in an arbitrage-free model of nominal and real bond yields." *Journal of Money, Credit and Banking*, 42(s1): 143–178.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. "Nominal rigidities and the dynamic effects of a shock to monetary policy." *Journal of Political Economy*, 113(1): 1–45.

Cieslak, Anna, and Pavol Povala. 2014. "Expecting the fed." Northwestern University Working Paper.

Cieslak, Anna, and Pavol Povala. 2015*a*. "Expected returns in treasury bonds." *Review of Financial Studies*, 28(10): 2859–2901.

Cieslak, Anna, and Pavol Povala. 2015*b*. "Information in the term structure of yield curve volatility." *Journal of Finance*, Forthcoming.

Cochrane, John H., and Monika Piazzesi. 2005. "Bond risk premia." *American Economic Review*, 95(1): 138–160.

Cochrane, John H., and Monika Piazzesi. 2008. "Decomposing the yield curve." Stanford University Working Paper.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A. Lochstoer. 2013. "Parameter learning in general equilibrium: The asset pricing implications." National Bureau of Economic Research Working Paper.

Coroneo, Laura, Domenico Giannone, and Michele Modugno. 2015. "Unspanned macroeconomic factors in the yield curve." *Journal of Business and Economic Statistics*.

Coroneo, Laura, Ken Nyholm, and Rositsa Vidova-Koleva. 2011. "How arbitrage-free is the Nelson-Siegel model?" *Journal of Empirical Finance*, 18(3): 393–407.

Cremers, K.J. Martijn. 2002. "Stock return predictability: A Bayesian model selection perspective." *Review of Financial Studies*, 15(4): 1223–1249.

D'Agostino, Antonello, Domenico Giannone, and Paolo Surico. 2006. "(Un)Predictability and macroeconomic stability." European Central Bank Working Paper Series 0605.

Dai, Qiang, and Kenneth J. Singleton. 2000. "Specification analysis of affine term structure models." *The Journal of Finance*, 55(5): 1943–1978.

Dai, Qiang, and Kenneth J. Singleton. 2003. "Term Structure Dynamics in Theory and Reality." *Review of Financial Studies*, 16(3): 631–678.

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

Della Corte, Pasquale, Lucio Sarno, and Daniel L. Thornton. 2008. "The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value." *Journal of Financial Economics*, 89(1): 158–174.

Den Haan, Wouter J., and Steven W. Sumner. 2004. "The comovement between real activity and prices in the G7." *European Economic Review*, 48(6): 1333–1347.

Dewachter, Hans, and Leonardo Iania. 2012. "An extended macro-finance model with financial factors." *Journal of Financial and Quantitative Analysis*, 46(06): 1893–1916.

Dewachter, Hans, and Marco Lyrio. 2008. "Learning, macroeconomic dynamics and the term structure of interest rates." In *Asset Prices and Monetary Policy*. 191–245. University of Chicago Press.

Dewachter, Hans, Leonardo Iania, and Marco Lyrio. 2014. "Information in the yield curve: A Macro-Finance approach." *Journal of Applied Econometrics*, 29(1): 42–64.

Diebold, Francis X., and Canlin Li. 2006. "Forecasting the term structure of government bond yields." *Journal of Econometrics*, 130(2): 337–364.

Diebold, Francis X., and Glenn D. Rudebusch. 2013. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach. The Econometric and Tinbergen Institutes Lectures*, Princeton University Press.

Diebold, Francis X., and Kamil Yilmaz. 2009. "Measuring financial asset return and volatility spillovers, with application to global equity markets." *The Economic Journal*, 119(534): 158–171.

Diebold, Francis X, and Kamil Yilmaz. 2014. "On the network topology of variance decompositions: Measuring the connectedness of financial firms." *Journal of Econometrics*, 182(1): 119–134.

Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13(3): 253–263.

Diebold, Francis X., Canlin Li, and Vivian Z. Yue. 2008. "Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach." *Journal of Econometrics*, 146(2): 351–363.

Diebold, Francis X., Glenn D. Rudebusch, and S. Borağan Aruoba. 2006. "The macroeconomy and the yield curve: A dynamic latent factor approach." *Journal of Econometrics*, 131(1): 309–338.

Duffee, Gregory R. 2002. "Term premia and interest rate forecasts in affine models." *Journal of Finance*, 57(1): 405–443.

Duffee, Gregory R. 2010. "Sharpe ratios in term structure models." Department of Economics, Johns Hopkins University Working Paper.

Duffee, Gregory R. 2011*a*. "Forecasting with the term structure: The role of no-arbitrage restrictions." Johns Hopkins University, Department of Economics Working Paper.

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

Duffee, Gregory R. 2013. "Bond pricing and the macroeconomy." In *Handbook of the Economics of Finance*. Vol. 2, Part B, , ed. George M. Constantinides, Milton Harris and Rene M. Stulz, 907–967. Elsevier.

Duffee, Gregory R. 2014. "Expected inflation and other determinants of Treasury yields." Johns Hopkins University, Department of Economics Working Paper.

Duffee, Gregory R, and Richard H Stanton. 2012. "Estimation of dynamic term structure models." *The Quarterly Journal of Finance*, 02(02): 1250008.

Duffie, Darrell. 2001. *Dynamic Asset Pricing Theory.* Princeton University Press.

Duffie, Darrell, and Rui Kan. 1996. "A yield-factor model of interest rates." *Mathematical Finance*, 6(4): 379–406.

Eickmeier, Sandra, Leonardo Gambacorta, and Boris Hofmann. 2014. "Understanding global liquidity." *European Economic Review*, 68: 1–18.

Elliott, Graham, and Allan Timmermann. 2008. "Economic forecasting." *Journal of Economic Literature*, 46(1): 3–56.

Estrella, Arturo, and Frederic S. Mishkin. 1998. "Predicting US recessions: Financial variables as leading indicators." *Review of Economics and Statistics*, 80(1): 45–61.

Evans, Charles L., and David A. Marshall. 2007. "Economic determinants of the nominal treasury yield curve." *Journal of Monetary Economics*, 54(7): 1986–2003.

Evans, George W., and Seppo Honkapohja. 2001. *Learning and Expectations in Macroeconomics.* Princeton University Press.

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

Feng, Ziding D., and Charles E. McCulloch. 1996. "Using bootstrap likelihood ratios in finite mixture models." *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(3): 609–617.

Fernández-Villaverde, Jesús, and Thomas J. Sargent Mark W. Watson Rubio-Ramírez, Juan F. 2007. "ABCs (and Ds) of Understanding VARs." *The American Economic Review*, 97(3): 1021–1026.

Feunou, Bruno, Jean-Sébastien Fontaine, Anh Le, and Christian Lundblad. 2014. "Term structure modeling when monetary policy is unconventional: A new approach." Bank of Canada Working Paper.

Filipović, Damir. 1999. "A note on the Nelson–Siegel family." *Mathematical Finance*, 9(4): 349–359.

Gagliardini, Patrick, Paolo Porchia, and Fabio Trojani. 2009. “Ambiguity aversion and the term structure of interest rates.” *Review of Financial Studies*, 22(10): 4157–4188.

Gargano, Antonio, Davide Pettenuzzo, and Allan G. Timmermann. 2014. “Bond return predictability: Economic value and links to the macroeconomy.” University of California, San Diego Working Paper.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang. 2007. “Portfolio selection with parameter and model uncertainty: A multi-prior approach.” *Review of Financial Studies*, 20(1): 41–81.

Geweke, John, and Gianni Amisano. 2010. “Comparing and evaluating Bayesian predictive distributions of asset returns.” *International Journal of Forecasting*, 26(2): 216–230.

Giacoletti, Marco, Kristoffer T. Laursen, and Kenneth J. Singleton. 2014. “Learning, dispersion of beliefs, and risk premiums in an aribitrage-free term structure model.” Stanford University Working Paper.

Gilboa, Itzhak, and David Schmeidler. 1989. “Maxmin expected utility with non-unique prior.” *Journal of Mathematical Economics*, 18(2): 141–153.

Gürkaynak, Refet S., and Jonathan H. Wright. 2012. “Macroeconomics and the term structure.” *Journal of Economic Literature*, 50(2): 331–367.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright. 2007. “The US Treasury yield curve: 1961 to the present.” *Journal of Monetary Economics*, 54(8): 2291–2304.

Hansen, Lars Peter. 2007. “Beliefs, doubts and learning: Valuing macroeconomic risk.” *American Economic Review*, 97(2): 1–30.

Hanson, Bruce E. 2002. “Tests for parameter instability in regressions with I(1) processes.” *Journal of Business and Economic Statistics*, 20(1): 45–59.

Hautsch, Nikolaus, and Fuyu Yang. 2012. “Bayesian inference in a stochastic volatility Nelson-Siegel model.” *Computational Statistics and Data Analysis*, 56(11): 3774–3792.

Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho. 2011. “What factors drive global stock returns?” *Review of Financial Studies*, 24(8): 2527–2574.

Hubrich, Kirstin, Antonello D’Agostino, et al. 2013. “Financial shocks and the macroeconomy: Heterogeneity and non-linearities.” European Central Bank ECB Occasional Paper Series 143.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. “Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software.” *PlOS ONE*, 9(6): e98679.

Johannes, Michael, Arthur Korteweg, and Nicholas Polson. 2013. “Sequential learning, predictability, and optimal portfolio returns.” *The Journal of Finance*, 69(2): 4.

Joslin, Scott, Anh Le, and Kenneth J. Singleton. 2013. “Gaussian macro-finance term structure models with lags.” *Journal of Financial Econometrics*, 11(4): 581–609.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu. 2011. “A new perspective on Gaussian dynamic term structure models.” *Review of Financial Studies*, 24(3): 926–970.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton. 2014. “Risk premiums in dynamic term structure models with unspanned macro risks.” *The Journal of Finance*, 69(3): 1197–1233.

Jotikasthira, Pab, Anh Le, and Christian T. Lundblad. 2015. "Why do term structures in different currencies comove?" *Journal of Financial Economics*, 115(1): 58–83.

Kim, Chang-Jin, and Charles R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* Vol. 1, the MIT Press.

Kim, Don H. 2009. "Challenges in Macro-Finance modeling." *Federal Reserve Bank of St. Louis Review*, 91(5, Part 2): 519–44.

Kim, Don H., and Jonathan H. Wright. 2005. "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates." Board of Governors of the Federal Reserve System Working Paper.

Kim, Don H., and Kenneth J. Singleton. 2012. "Term structure models and the zero bound: An empirical investigation of Japanese yields." *Journal of Econometrics*, 170(1): 32–49.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit.* Hougton Mifflin Company.

Koop, Gary, and Dimitris Korobilis. 2009. "Bayesian multivariate time series methods for empirical macroeconomics." *Foundations and Trends® in Econometrics*, 3(4): 267–358.

Koop, Gary, and Dimitris Korobilis. 2012. "Forecasting inflation using Dynamic Model Averaging." *International Economic Review*, 53(3): 867–886.

Koop, Gary, and Dimitris Korobilis. 2013. "Large time-varying parameter VARs." *Journal of Econometrics*, 177(2): 185–198.

Koop, Gary, and Dimitris Korobilis. 2014. "A new index of financial conditions." *European Economic Review*, 71(0): 101 – 116.

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter. 1996. "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, 74(1): 119–147.

Koopman, Siem Jan, Max I.P. Mallee, and Michel Van der Wel. 2010. "Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters." *Journal of Business and Economic Statistics*, 28(3): 329–343.

Kose, M. Ayhan, Christopher Otrok, and Charles H. Whiteman. 2003. "International business cycles: World, region, and country-specific factors." *American Economic Review*, 93(4): 1216–1239.

Kozicki, Sharon, and Peter A. Tinsley. 2001. "Shifting endpoints in the term structure of interest rates." *Journal of Monetary Economics*, 47(3): 613–652.

Krippner, Leo. 2015. "A theoretical foundation for the Nelson-Siegel class of yield curve models." *Journal of Applied Econometrics*, 30(1): 97–118.

Kumar, Alok, and Charles Lee. 2006. "Retail investor sentiment and return comovements." *The Journal of Finance*, 61(5): 2451–2486.

Kurmann, André, and Christopher Otrok. 2013. "News shocks and the slope of the term structure of interest rates." *American Economic Review*, 103(6): 2612–32.

Kurz, Mordecai. 1994. "On rational belief equilibria." *Economic Theory*, 4(6): 859–876.

Laubach, Thomas, Robert J. Tetlow, and John C. Williams. 2007. "Learning and the role of macroeconomic factors in the term structure of interest rates." Board of Governors of the Federal Reserve System Working Paper.

Lee, Bong-Soo. 1998. "Permanent, temporary, and non-fundamental components of stock prices." *Journal of Financial and Quantitative Analysis*, 33(01): 1–32.

Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang. 2013. "Fiscal foresight and information flows." *Econometrica*, 81(3): 1115–1145.

Litterman, Robert B., and Jose Scheinkman. 1991. "Common factors affecting bond returns." *The Journal of Fixed Income*, 1(1): 54–61.

Litterman, Robert B., José Scheinkman, and Laurence Weiss. 1991. "Volatility and the yield curve." *The Journal of Fixed Income*, 1(1): 49–53.

Liu, Jun, Jun Pan, and Tan Wang. 2005. "An equilibrium model of rare-event premia and its implication for option smirks." *Review of Financial Studies*, 18(1): 131–164.

Ludvigson, Sydney C. 2004. "Consumer confidence and consumer spending." *The Journal of Economic Perspectives*, 18(2): 29–50.

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

Markowitz, Harry. 1952. "Portfolio Selection." *The Journal of Finance*, 7(1): 77–91.

McCulloch, J. Huston. 2007. "The Kalman foundations of adaptive least squares, with application to US inflation." Ohio State University Working Paper.

Moench, Emanuel. 2008. "Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach." *Journal of Econometrics*, 146(1): 26–43.

Moench, Emanuel. 2012. "Term structure surprises: The predictive content of curvature, level, and slope." *Journal of Applied Econometrics*, 27(4): 574–602.

Moench, Emanuel, Serena Ng, and Simon Potter. 2013. "Dynamic hierarchical factor models." *Review of Economics and Statistics*, 95(5): 1811–1817.

Neftci, Salih. 2004. *Principles of Financial Engineering.* Academic Press Inc.

Nelson, Charles R., and Andrew F. Siegel. 1987. "Parsimonious modeling of yield curves." *Journal of Business*, 60(4): 473–489.

Novy-Marx, Robert. 2014. "Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars." *Journal of Financial Economics*, 112(2): 137–146.

Orphanides, Athanasios, and Min Wei. 2012. "Evolving macroeconomic perceptions and the term structure of interest rates." *Journal of Economic Dynamics and Control*, 36(2): 239–254.

Pástor, L'uboš, and Robert F Stambaugh. 1999. "Costs of equity capital and model mispricing." *The Journal of Finance*, 54(1): 67–121.

Pástor, L'uboš, and Robert F Stambaugh. 2000. "Comparing asset pricing models: an investment perspective." *Journal of Financial Economics*, 56(3): 335–381.

Pettenuzzo, Davide, Allan Timmermann, and Rossen Valkanov. 2014. "Forecasting stock returns under economic constraints." *Journal of Financial Economics*, 114(3): 517–553.

Piazzesi, Monika. 2010. "Affine term structure models." *Handbook of Financial Econometrics*, 1: 691–766.

Piazzesi, Monika, and Martin Schneider. 2007. "Equilibrium yield curves." In *NBER Macroeconomics Annual 2006, Volume 21*. 389–472. MIT Press.

Pooter, Michiel De. 2007. "Examining the Nelson-Siegel class of term structure models." Tinbergen Institute Discussion Paper.

Prado, R., and M. West. 2010. *Time Series: Modeling, Computation, and Inference. Chapman & Hall/CRC Texts in Statistical Science*, Taylor & Francis.

Raftery, Adrian E., Miroslav Kárnỳ, and Pavel Ettler. 2010. "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill." *Technometrics*, 52(1): 52–66.

Rapach, David E., Jack K. Strauss, and Guofu Zhou. 2009. "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy." *Review of Financial Studies*, 23(2): 821–862.

Rossi, Barbara. 2005. "Optimal tests for nested model selection with underlying parameter instability." *Econometric theory*, 21(05): 962–990.

Sangvinatsos, Antonios, and Jessica A. Wachter. 2005. "Does the failure of the expectations hypothesis matter for long-term investors?" *The Journal of Finance*, 60(1): 179–230.

Sargent, Thomas J. 2002. *The Conquest of American Inflation.* Princeton University Press.

Sarno, Lucio, Daniel L. Thornton, and Giorgio Valente. 2007. "The empirical failure of the expectations hypothesis of the term structure of bond yields." *Journal of Financial and Quantitative Analysis*, 42(01): 81–100.

Sarno, Lucio, Paul Schneider, and Christian Wagner. 2014. "The economic value of predicting bond risk premia: Can anything beat the expectations hypothesis." Cass Business School and Centre for Economic Policy Research (CEPR) Working Paper.

Sharpe, William F. 1970. *Portfolio Theory and Capital Markets.* Vol. 217, McGraw-Hill New York.

Siegel, Andrew F., and Charles R. Nelson. 1988. "Long-term behavior of yield curves." *Journal of Financial and Quantitative Analysis*, 23(01): 105–110.

Stock, James H., and Mark W. Watson. 2002. "Forecasting using principal components from a large number of predictors." *Journal of the American Statistical Association*, 97(460): 1167–1179.

Tang, Huarong, and Yihong Xia. 2007. "An international examination of affine term structure models and the expectations hypothesis." *Journal of Financial and Quantitative Analysis*, 42(01): 41–80.

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

Timmermann, Allan G. 1993. "How learning in financial markets generates excess volatility and predictability in stock prices." *The Quarterly Journal of Economics*, 108(4): 1135–1145.

Timmermann, Allan G. 1996. “Excess volatility and predictability of stock prices in autoregressive dividend models with learning.” *The Review of Economic Studies*, 63(4): 523–557.

Ulrich, Maxim. 2013. “Inflation ambiguity and the term structure of us government bonds.” *Journal of Monetary Economics*, 60(2): 295–309.

Uppal, Raman, and Tan Wang. 2003. “Model misspecification and underdiversification.” *The Journal of Finance*, 58(6): 2465–2486.

Van Dijk, Dick, Siem Jan Koopman, Michel Van der Wel, and Jonathan H. Wright. 2014. “Forecasting interest rates with shifting endpoints.” *Journal of Applied Econometrics*, 29(5): 693–712.

Wan, Eric A., and Rudolph Van Der Merwe. 2000. “The unscented Kalman filter for nonlinear estimation.” In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*. 153–158. IEEE.

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

West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag New York, Inc.

Wheelock, David C., and Mark E. Wohar. 2009. “Can the term spread predict output growth and recessions? A survey of the literature.” *Federal Reserve Bank of St. Louis Review*, 91(Part 1): 419–440.

Wright, Jonathan H. 2006. “The yield curve and predicting recessions.” Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board Working Paper 07.

Wright, Jonathan H. 2011. “Term premia and inflation uncertainty: Empirical evidence from an international panel dataset.” *American Economic Review*, 101(4): 1514–1534.

Appendices

B.1 Discussion about Model Specification

B.1.1 Macro-Spanning Condition

To test whether macro variables can be spanned by bond yields in our sample period, we regress inflation and industrial production on principal components (PCs) of bond yields. Table 10 shows macro variables are weakly spanned by PCs, which parallels the finding in Bauer and Rudebusch (2015) that macro variables may not be spanned by lower-order PCs. This is because the principal component method only considers cross-section variance, see Stock and Watson (2002), and Bauer and Rudebusch (2015) suggest high-order PCs that are useful in spanning macro factors are likely to be contaminated by measurement errors.

Table 10: *Economic Measure Regressions on Bond Yield Factors*

	CPI			IP		
	2 PCs	3 PCs	5 PCs	2 PCs	3 PCs	5 PCs
Global	**8.24%**	8.56%	28.63%	**7.20%**	17.05%	16.62%
US	9.88%	13.26%	38.84%	18.37%	22.07%	27.16%
UK	3.12%	2.69%	18.94%	23.12%	23.22%	56.08%
JP	-0.50%	0.07%	3.75%	3.91%	4.47%	9.28%
GER	13.03%	12.72%	32.48%	8.99%	8.63%	19.05%
FRA	2.14%	2.30%	6.74%	0.41%	0.79%	16.37%
CAN	18.37%	19.46%	38.25%	22.08%	22.15%	31.73%
ITA	17.33%	28.69%	29.66%	9.22%	8.86%	28.14%

Notes: This table reports the Adjusted R_2 of regressions in which CPI inflation and industrial production growth rate (year on year) are regressed on different numbers of principal components (PCs) of bond yields. The sample is from 1994:12 to 2014:03 at monthly frequency. The global variables are G7 aggregates from OECD database.

Therefore, we adopt the unspanned restrictions advocated by the data for parsimony, and Bauer and Rudebusch (2015) suggest that spanned and unspanned models deliver similar results. It is worth highlighting the robustness of Moench, Ng and Potter (2013): Unlike principal components, this method identifies factors by allowing for not only cross-sectional variance but also time series properties. This also means, even in the extreme case that unspanned restrictions are not necessary, the identified factors will cater to the true dynamics and hence mitigate specification errors. The potential loss caused the parsimonious unspanned setup, if any, should be economically insignificant.

Note that unspanned restrictions do not violate Taylor-type policy rules. To see this, we write down the restrictions about macro variables M_t following Bauer and Rudebusch (2015):

$$M_t = \gamma_0 + \gamma_P P_t^L + OM_t,$$

where OM_t captures the orthogonal macroeconomic variation not captured by lower-order PCs P_t^L. For convenience, assuming M_t, P_t have the same dimension and γ_P is invertible, then the short rate r_t is a linear function of PCs and hence a linear function of M_t:

$$r_t = \beta P_t^L = C(\gamma_0, \gamma_P, OM_t) + \beta \gamma_p^{-1} M_t,$$

where C is a function of $(\gamma_0, \gamma_P, OM_t)$. It is clear the short rate is a linear function of macro variables. From the equation we see that the time-varying unspanned variance in OM_t can potentially contaminate the explanatory power of M_t.

The *macro spanning condition* should not be confused with the issue whether bond yields are significantly driven by macro factors. That is, even we assume macro factors are fully spanned by bond yields, macro factors do not necessarily have higher explanatory power for yields. The *macro spanning condition* is only about whether bond factors include all information of macro variables that can be used to estimate term premia, and term premia is always a linear function of macro factors in a macro-finance model, no matter the factors are spanned or not.[24] A separate but related questions is, how much of the variance of bond yields can be explained by macro factors and why. This question is what we are trying to answer in this paper, and our results are considered robust with the identification strategy proposed by Moench, Ng and Potter (2013), as the pricing factors are identified allowing for time-series information of global macro fundamentals.

B.1.2 Cross-Sectional Restrictions

In this paper, we do not impose no-arbitrage constraints in our model as the constraints are silent about identifying the latent factors and shocks. Duffee (2013) suggests Nelson-Siegel restrictions are nearly equivalent to no-arbitrage in characterizing the cross section of interest rate term structure. Joslin, Le and Singleton (2013) show that Gaussian no-arbitrage macro-finance models are close to factor-VAR models when risk premia dynamics are not constrained. Duffee (2014) also indicates that the no-arbitrage restrictions are unimportant if a model aims to pin down physical dynamics. Since our focus here is not on the structure of risk premia dynamics, we choose to impose no such restrictions to avoid potential misspecification. The potential drawback of no-arbitrage models is that it imposes very strong restrictions on the dynamics of risk prices, in order to 1) ensure no-arbitrage consumption and 2) identify the model with flat likelihood. Kim and Singleton (2012) and Jotikasthira, Le and Lundblad (2015) indicate the no-arbitrage framework may generate implausibly term premiums in the financial crisis. Instead, we impose Nelson-Siegel restrictions here, which provide a parsimonious structure and satisfactory performance in cross-sectional fittings of term structure.

[24] Macro spanning, by construction, means macro factors are a subset of pricing factors, and therefore pricing factors have all information of macro factors. This intuition has been discussed formally in Duffee (2013).

B.2 Econometric Methods

In this paper we propose a novel approach which extends the hierarchical factor model of Moench, Ng and Potter (2013) by augmenting the model with macro factors. We apply the NS restrictions similar to Diebold, Li and Yue (2008) for the yield factor identification. The estimation of our model is in one step, which should provide more accurate estimates when compared to other multi-step estimations. We call the new model 'Fundamentals-Augmented Hierarchical Factor Model' (FAHFaM).

Our proposed hierarchical model has three levels of factor dynamics, but we only focus on the global level that is augmented with global macro factors. At the global level, the dynamics of the global yield factors can be regarded as an unrestricted Factor-Augmented Vector Autoregressive (FAVAR) system. We conduct the analysis in two steps. The first step is to extract the latent global yield factors, using the proposed 'Fundamentals-Augmented Hierarchical Dynamic Factor Model'. The second step is to directly use the estimation results of FAVAR at the global level to identify the shocks of interest.

B.2.1 Fundamentals-Augmented Hierarchical Factor Model

To extract the latent factors, a principal component method is commonly utilized. Bai and Ng (2006) have shown that the estimated factors from the principal components method can be treated as though they are observed, if $\sqrt{T}/N \to \infty$ as $T, N \to \infty$. However, the method of principal components is not well suited for the present analysis, because the number of series available[25] is much smaller than the large dimensions that the principal component method typically requires. Accordingly, the FAHFaM is proposed to extract the latent global factors.

B.2.1.1 A Three-Level Hierarchical Factor Model

Following the framework developed by Moench, Ng and Potter (2013), a three-level model is considered here. Level one is the national level, which describes how national yield factors drive the yields at different maturities. Level two is the global-national level, illustrating how the global yield factors govern the national yield factors. Level three displays the autoregressive dynamics of the global factors.

Firstly, we treat a block (identified as b) as one of the seven countries, so $b = 1, 2, ..., B$ where $B = 7$. At the national level, the bond yield data for a specific country are stacked in the vector X_{bt}, and the dynamic representation is given by

$$X_{b,t} = \Lambda_b^F F_{b,t} + e_{b,t}^X, \tag{4}$$

[25] There are only seven countries so $N = 7$.

where $X_{b,t}$ is an $N_b \times 1$ vector of yields of country b at different maturities, $F_{b,t}$ is a $k_b \times 1$ vector of latent common yield factors at national level, Λ_b^F is an $N_b \times k_b$ coefficient matrix and $e_{b,t}^X$ is the vector of idiosyncratic components. Note that in our model $N_b = 11$ and $k_b = 2$ for $b = 1,2,...,B$; in other words, for each country, we use yield data of 11 different maturities and assume that 2 factors can explain most of the yield variance.

Stacking up $F_{b,t}$ across seven countries produces a $K^F \times 1$ vector F_t.[26] At the global-national level, it is assumed that

$$F_t = \Lambda^G G_t + e_t^F, \tag{5}$$

where K^G global common factors are collected into the vector G_t, Λ^G is a $K^F \times K^G$ coefficient matrix and e_t^F are country-specific components at the global-national level.

The dynamics of the global factors G_t are described at level three:

$$G_t = \Psi^G G_{t-1} + \varepsilon_t^G, \tag{6}$$

where Ψ^G is the coefficient matrix and the innovations $\varepsilon_t^G \sim N\big(0, \Sigma^G\big)$[27].

The model is completed by specifying the dynamics of idiosyncratic and country-specific components $e_{b,t}^X$ and e_t^F.

$$e_{b,t}^X = \Psi_b^X e_{b,t-1}^X + \varepsilon_{b,t}^X, \tag{7}$$

$$e_t^F = \Psi^F e_{t-1}^F + \varepsilon_t^F, \tag{8}$$

where Ψ_b^X is an $N_b \times N_b$ diagonal coefficient matrix, Ψ^F is a $K^F \times K^F$ diagonal coefficient matrix, the innovations $\varepsilon_{b,t}^X \sim N\big(0, \Sigma_b^X\big)$ and $\varepsilon_t^F \sim N\big(0, \Sigma^F\big)$.[28]

B.2.1.2 An Extension with Macro Factor Augmentation

Assuming at level three, i.e. the level that describes the global factor dynamics, the factor dynamics are augmented with Macro information. So the Equation (6) can be rewritten as

$$\begin{bmatrix} G_t \\ M_t \end{bmatrix} = \psi^G \begin{bmatrix} G_{t-1} \\ M_{t-1} \end{bmatrix} + u_t, \tag{9}$$

$$u_t \sim N\big(\mathbf{0}, \Sigma^G\big),$$

where Σ^G is the variance-covariance matrix of u_t. The evolution of the global factors displayed here uses only one lag here for simplicity; in practice, more lags can be used to estimate the factor

[26] $K^F = \sum_{b=1}^B k_b$ and $F_t = \begin{pmatrix} F_{1,t} & F_{2,t} & \dots & F_{B,t} \end{pmatrix}'$

[27] $\Sigma^G = \mathrm{diag}\big((\sigma_1^G)^2, ..., (\sigma_{K^G}^G)^2\big)$.

[28] $\Sigma_b^X = \mathrm{diag}\big((\sigma_{b,1}^X)^2, ..., (\sigma_{b,N_b}^X)^2\big)$ and $\Sigma^F = \mathrm{diag}\big((\sigma_1^F)^2, ..., (\sigma_{K^F}^F)^2\big)$.

dynamics. The Equation (9) is indeed a factor-augmented vector autoregressive (FAVAR) system. The estimates from this system will be used for the identification of shocks for the structural analysis.

B.2.1.3 Estimation via Gibbs Sampling

Before we proceed with the estimation scheme, the parameters needed to be estimated are summarized for better illustration. Collect $\{\Lambda_1^F,...,\Lambda_B^F\}$ and Λ^G into $\mathbf{\Lambda}$, $\{\Psi_1^X,...,\Psi_B^X\}$, Ψ^F and Ψ^G into $\mathbf{\Psi}$, and $\{\Sigma_1^X,...,\Sigma_B^X\}$, Σ^F, Σ^G into $\mathbf{\Sigma}$. To sum up, the parameters we need to estimate are $\mathbf{\Lambda}$, $\mathbf{\Psi}$ and $\mathbf{\Sigma}$.

A Bayesian method, i.e., Markov Chain Monte Carlo (MCMC), is used to estimate the model. A simple extension of the algorithm in Carter and Kohn (1994) is proposed here. Based on the observed values of M_t, and the initial values of $\{F_{b,t}\}$ and G_t from the method of principal components, for each iteration we construct the Gibbs sampler in the following steps:

1. Draw G_t, conditional on F_t, $\mathbf{\Lambda}$, $\mathbf{\Psi}$ and $\mathbf{\Sigma}$.
2. Draw Ψ^G, conditional on Σ^G, G_t and M_t.
3. Draw Σ^G, conditional on Ψ^G, G_t and M_t.
4. Draw Λ^G, conditional on G_t and F_t.
5. For each b, draw $F_{b,t}$, conditional on $\mathbf{\Lambda}$, $\mathbf{\Psi}$, $\mathbf{\Sigma}$ and G_t.
6. For each b, draw b_{th} elements of Ψ^F and Σ^F, conditional on G_t and F_t.
7. For each b, draw the Λ_b^F, Ψ_b^X and Σ_b^X, conditional on F_t and $X_{b,t}$.

Similar to Diebold, Li and Yue (2008) and Moench, Ng and Potter (2013), the elements of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ are set to have normal priors, and $\mathbf{\Sigma}$ follow inverse gamma priors. Given the conjugacy, the posterior distributions are not difficult to compute. Regarding the factors G_t and F_t, we follow Carter and Kohn (1994) and Kim and Nelson (1999) to run the Kalman filter forward to obtain the estimates in period T and then proceed backward to generate draws for $t = T-1,...,1$. It is worth noting that, if we impose hard restrictions on Λ^G and Λ_b^F, then there is no need to draw these parameters in the above Gibbs sampling.

B.2.2 Nelson-Siegel Restrictions

Following Diebold, Li and Yue (2008), we can use two factors to summarize most of the information in the term structure of interest rates. As we show in the Section 3.3.1, two factors have accounted for around 99% of the bond yield variance across all countries.

The below Equation (10) describes how restrictions are imposed; the restrictions used in our hierarchical factor model are in fact fixing the loading of the factors. Let $y_t(\tau)$ denote yields at maturity τ, then the factor model for a single country we use is of the form

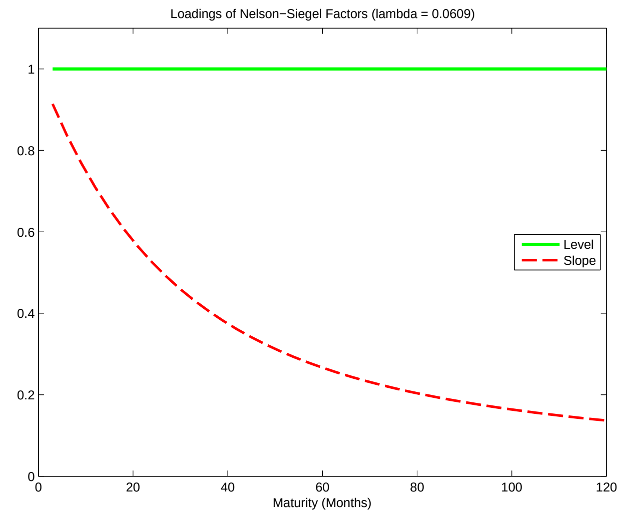
$$y_t(\tau) = L_t^{NS} + \frac{1 - e^{-\tau\lambda}}{\tau\lambda} S_t^{NS} + \varepsilon_t(\tau), \tag{10}$$

where L_t^{NS} is the "Level" factor, S_t^{NS} is the "Slope" factor, and ε_t is the error term. Additionally, λ in the exponential functions controls the shapes of loadings for the NS factors; following Diebold and Li (2006) and Bianchi, Mumtaz and Surico (2009), we set $\lambda = 0.0609$.[29]

The interpretations of Nelson-Siegel factors are of empirical significance. The Nelson-Siegel Level factor L_t^{NS} is identified as the factor that is loaded evenly by the yields of all maturities. The Slope factor S_t^{NS} denotes the spread between the yields of a short- and a long-term bond, and its movements are captured by putting more weights on the yields with shorter maturities.

The following Figure 6 depicts the shapes of the loadings of the NS factors. In our model estimation, we fixed the Λ_b^F in Equation (4) by the NS loadings. We further set the Λ^G in Equation (5) to a diagonal matrix to identify the global factors, and the intuition behind is that the country-level Level (Slope) factor is only driven by the global Level (Slope) factor.

Figure 6: *Loadings of Nelson-Siegel Factors*



Notes: The solid green line and red dashed line are the loadings for Level and Slope factors, respectively ($\lambda = 0.0609$). The horizontal axis shows the maturities of bonds, and the unit is month.

[29] Alternatively, we can select the value of λ from a grid of reasonable values by comparing the goodness of fit. However, if we do not specify the factor dynamics and fit the Nelson-Siegel model in a static way, the selection may not be optimal. Also we choose a single value of λ for all the countries, as Nelson and Siegel (1987) indicate that there is little gain in practice by fitting λ individually. Therefore, we set $\lambda = 0.0609$ to fix the ideas because 1) this value is the mostly used in the related literature so revealing the dynamics the associate latent factors is more desirable, and 2) using this value we have a relatively better fit of the 'global short rate factor'. To ensure the robustness, we also try a grid of reasonable values; we find the results are qualitatively similar and hence our findings are robust to the selection of λ.

B.2.3 Decomposition of Variance Driven by Global Factors

Recall Equation (6) that describes the dynamics of the global factors G_t at level three in Section B.2.1:

$$G_t = \Psi^G G_{t-1} + \varepsilon_t^G,$$

We can rewrite this as an implied Wold MA(∞) representation:

$$G_t = \sum_{i=0}^{\infty} \psi_i \mu_{t-i}, \tag{11}$$

where μ_t are the orthogonal innovations and Cholesky decomposition is needed to take into account the contemporaneous correlation of the shocks.

With simple algebra, we can write the bond yield co-movements driven by the global factors X_t^G as the following equation:

$$X_t^G = B \sum_{i=0}^{\infty} \psi_i \mu_{t-i}, \tag{12}$$

where B is the product of the loadings Λ^F (in Equation 4) and Λ^G (in Equation 5). The impulse response at time $t+h$ is therefore:

$$X_{t+h}^G = B \sum_{i=0}^{\infty} \psi_i \mu_{t+h-i}. \tag{13}$$

It is easy to have the error of the optimal h-step ahead forecast at time t:

$$X_{t+h}^G - \hat{X}_{t+h|t}^G = B \sum_{i=0}^{h-1} \psi_i \mu_{t+h-i}. \tag{14}$$

The mean squared error of X_{t+h}^G is given by

$$\mathrm{MSE}(X_{t+h}^G) = \mathrm{diag}\big(B(\sum_{i=0}^{h-1} \psi_i \psi_i')B'\big). \tag{15}$$

Therefore, the contribution of the kth factor to the MSE of the h-step ahead forecast of the yield at the jth maturity is

$$\Omega_{jk,h} = \sum_{i=0}^{h-1} R_{jk,i}^2 / \mathrm{MSE}(X_{t+h}^G), \tag{16}$$

where $R_{jk,i}$ is the element in row j, column k of $R_i = B\psi_i$.

B.2.3.1 Decomposition of Policy Channel and Risk Compensation Channel

The policy channel is consistent with the 'Expectation Hypothesis' (EH). The EH consistent long yield is given by

$$y_t(\tau)^{EH} = \frac{1}{\tau} \sum_{i=0}^{\tau-1} \mathrm{E}_t y_{t+i}(1), \tag{17}$$

where $y_t(\tau)$ is the element of yield data X_t at maturity τ. That is to say, the EH consistent long yield is equal to the average of expected short yields $\mathrm{E}_t y_{t+i}(1)$. If we only focus on the part driven by global factors, then after some iterations, the above equation can be written as

$$y_t(\tau)^{EH} = \frac{1}{\tau}B(I + \Psi^G + \Psi^{G^2} + ... + \Psi^{G^{\tau-1}}) \sum_{i=0}^{\infty} \psi_i \mu_{t-i}. \tag{18}$$

The term premia (risk compensation channel) is given by

$$TP_t(\tau) = y_t(\tau) - y_t(\tau)^{EH}. \tag{19}$$

In other words, the term premia is the difference between the long yield and the EH consistent long yield. We can use similar transformations as in Equations (13) and (16) to compute the impulse response and variance decomposition of the above two channels.

B.2.4 Spillover Table and Generalized Variance Decomposition

The generalized variance decomposition (GVD) framework of Koop, Pesaran and Potter (1996) produces variance decompositions invariant to ordering. The GVD approach accounts for correlated shocks using the historically-observed error distribution, under a normality assumption. The GVD matrix has entries

$$\delta_{ij} = \frac{\delta_{jj}^{-1} \sum_{h=0}^{H-1} (e_i' A_h \Sigma e_j)^2}{\sum_{h=0}^{H-1} (e_i' A_h \Sigma A_h' e_i)} \tag{20}$$

where δ_{jj} is the variance of VAR shock ε_j, Σ is the covariance matrix of VAR shocks, A_h are MA(∞) coefficient matrices and e_j is a selection vector with jth element unity and zeros elsewhere. It means that shocks to variable j are responsible for $100 \times \delta_{ij}$ percent of the H-step-ahead forecast error variance in variable i.

Because shocks are not necessarily orthogonal in the GVD environment, sums of forecast error variance contributions, i.e. row sums in GVD matrices, are not necessarily unity. Therefore, the (i, j)-th entry in the spillover table is given by $100 \times \widetilde{\delta}_{ij} = 100 \times \frac{\delta_{ij}}{\sum_{j=1}^{N} \delta_{ij}}$, where N is the number of shocks. The Spillover Index is calculated from

$$SOI = \frac{\sum\limits_{\substack{i,j=1 \\ i \neq j}}^{N} \widetilde{\delta}_{ij}}{\sum_{i,j=1}^{N} \widetilde{\delta}_{ij}}. \tag{21}$$

B.3 Data Appendix

Table 11: *List of Financial Condition Indexes*

Series ID	Description
STLFSI	St. Louis Fed Financial Stress Index [1]
KCFSI	Kansas City Financial Stress Index [1]
ANFCI	Chicago Fed Adjusted National Financial Conditions Index [1]
CFSI	Cleveland Financial Stress Index [1]
VIX	CBOE S&P Volatility Index [1]
BFCIUS	Bloomberg United States Financial Conditions Index [1]
BFCIEU	Bloomberg Euro-Zone Financial Conditions Index [1]
GFSI	BofA Merrill Lynch Global Financial Stress Index [1]
EASSF	Euro Area Systemic Stress Indicator Financial Intermediary [1]
WJF	Westpac Japan Financial Stress Index [1]
GSF	Goldman Sachs Financial Index [1]
BCF	Bank of Canada Financial Conditions Index [1]

Notes:
1. In square brackets [·] we have a code for data transformations used in this data set: [1] means original series is used. The series are not seasonally adjusted.
2. Data are attained from Bloomberg, spanning from Jan. 1990 to Mar. 2014. The data may be unbalanced. The first five series can also be attained from St. Louis Federal Reserve Economic Data (http://research.stlouisfed.org/).

Table 12: *List of Yields*

Series ID	Description
ITA	Italy Sovereign (IYC 40) Zero Coupon Yields [1]
CAN	Canada Sovereign (IYC 7) Zero Coupon Yields [1]
FRA	France Sovereign (IYC 14) Zero Coupon Yields [1]
GER	German Sovereign (IYC 16) Zero Coupon Yields [1]
JP	Japan Sovereign (IYC 18) Zero Coupon Yields [1]
UK	United Kingdom (IYC 22) Zero Coupon Yields [1]
US	Treasury Actives (IYC 25) Zero Coupon Yields [1]

Notes:
1. In square brackets $[\cdot]$ we have a code for data transformations used in this data set: $[1]$ means original series is used. The series are not seasonally adjusted.
2. Data are attained from Bloomberg, spanning from Dec. 1994 to Mar. 2014. The yields are of the following 11 maturities: 3 months, 6 months, 1 year, 2 years, 3 years, 4 years, 5 years, 6 years, 7 years, 8 years and 10 years.
3. The zero-coupon yields are calculated step-by-step using the discount factors that are derived from standard bootstrapping, given the set of coupon bonds, bills, swaps or a combination of these instruments. A minimum of four instruments at different tenors are required for each yield curve. The bootstrapping is similar to the Unsmoothed Fama-Bliss method, see Fama and Bliss (1987).

Table 13: *List of Real Activity Indicators*

Series ID	Description
IMFIPUS	IMF US Industrial Production SA [5]
IMFIPUK	IMF UK Industrial Production SA [5]
IMFIPJP	IMF Japan Industrial Production SA [5]
IMFIPGER	IMF Germany Industrial Production SA [5]
IMFIPFR	IMF France Industrial Production SA [5]
IMFIPITA	IMF Italy Industrial Production SA [5]
IMFIPCAN	IMF Canada Industrial Production SA [5]

Notes:
1. In square brackets $[\cdot]$ we have a code for data transformations used in this data set: $[5]$ means log first-order difference (annually growth rate) is used.
2. Data are attained from Bloomberg, spanning from Jan. 1990 to Mar. 2014. The data may be unbalanced.

Table 14: *List of CPI and Policy Rates*

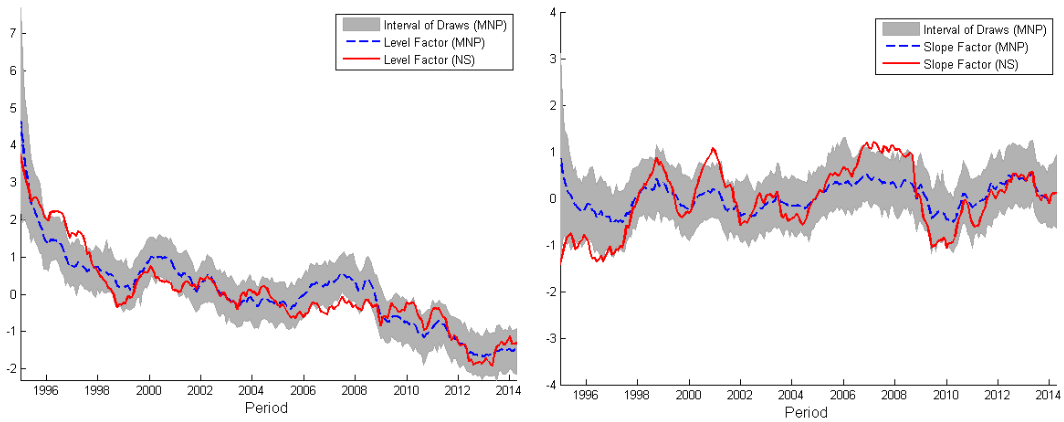
Series ID	Description
IMFCPIUS	IMF US CPI % Change in Percent per Annu [1]
IMFCPIUK	IMF UK CPI % Change in Percent per Annu [1]
IMFCPIJP	IMF Japan CPI % Change in Percent per Annu [1]
IMFCPIGER	IMF Germany CPI % Change in Percent per Annu [1]
IMFCPIFR	IMF France CPI % Change in Percent per Annu [1]
IMFCPIITA	IMF Italy CPI % Change in Percent per Annu [1]
IMFCPICAN	IMF Canada CPI % Change in Percent per Annu [1]
IMFFUNDUS	IMF US Federal Funds Rate in Percent per Annu [5]
IMFFUNDUK	IMF UK Bank of England Official Bank Rate [5]
IMFFUNDJP	IMF Japan Official Rate in Percent per Annu [5]
IMFFUNDCAN	IMF Canada Official Rate in Percent per Annu [5]
IMFFUNDEU	IMF Euro Area Official Rate in Percent per Annu [5]

Notes:
1. In square brackets $[\cdot]$ we have a code for data transformations used in this data set: $[1]$ means original series is used. The series are all seasonally adjusted; $[5]$ means log first-order difference (annually) is used.
2. Data are attained from Bloomberg, spanning from Jan. 1990 to Mar. 2014. The data may be unbalanced.

B.4 Additional Results

B.4.1 Comparison of Factor Indentification Schemes

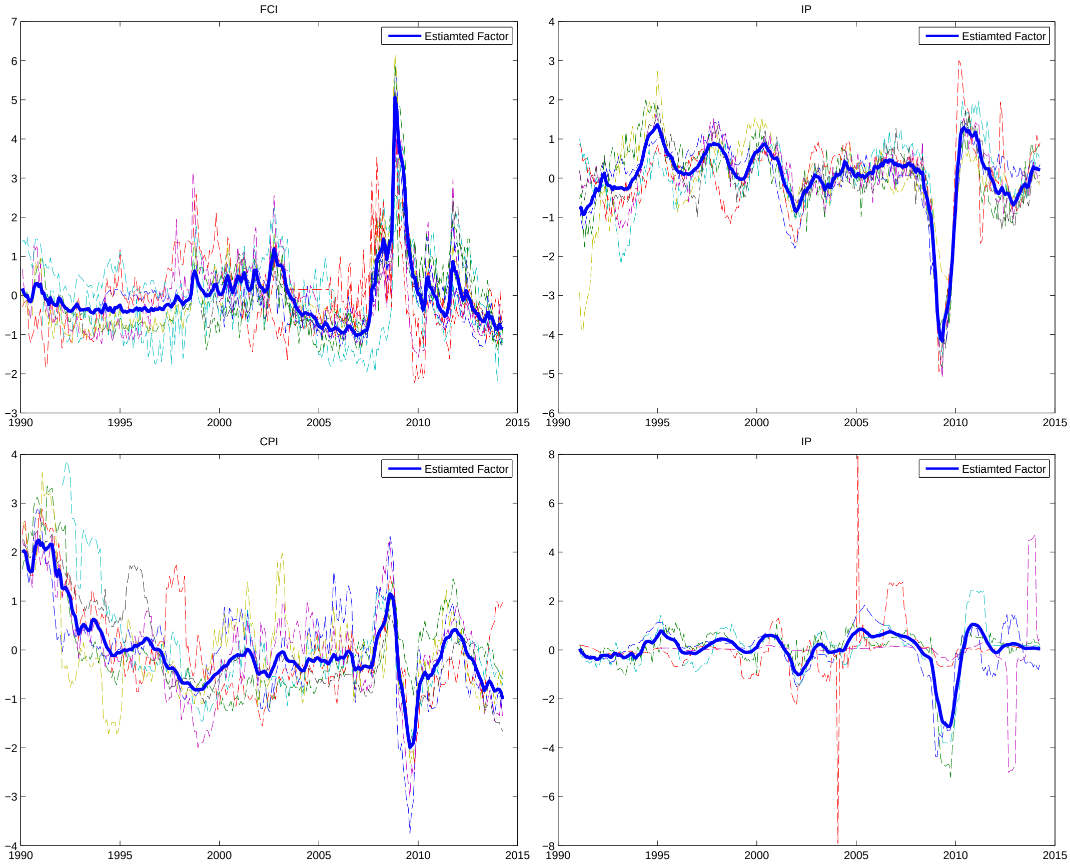
Figure 7: *Identified Factors from Different Schemes (MNP vs. NS)*



Notes:
1. In the above two charts, the factors identified by the scheme of Moench, Ng and Potter (2013) are plotted against the factors identified by the NS scheme of Diebold, Li and Yue (2008). To better serve the comparison purpose, the factors are extracted from a less complicated system without a macro factor augmentation.
2. The upper chart shows the Level factors, while the lower chart displays the Slope factor. The dashed blue lines are the median values of MNP identified factors and the gray areas cover all the draws from the MCMC estimation. The solid red lines are the median values of NS identified factors.

B.4.2 Global Macro Factors

Figure 8: *Estimated Global Macro Factors*



Notes:

1. In the above charts, the thick blue lines are the global macro factors, which are estimated using the method proposed by Koop and Korobilis (2014). The Matlab code can be obtained in website https://sites.google.com/site/dimitriskorobilis/matlab/. The other thin lines with different colors are the standardized series for the estimation.

2. From top left clock-wise we have global factors of financial condition indexes, real activity, policy rates and inflation. The data used for the factor estimation are described in Appendix B.3, spanning from Jan. 1990 to Mar. 2014.

Table 15: *Correlations between the National Series and Global Factors*

FCI	STLFSI	KCFSI	ANFCI	CFSI	VIX	BFCIUS	BFCIEU	GFSI	EASSF	WJF	GSF	BCF
Correlation	0.945	0.952	0.568	0.695	0.845	0.935	0.848	0.866	0.701	0.528	0.671	0.814

IP	IMFIPUS	IMFIPUK	IMFIPJP	IMFIPGER	IMFIPFR	IMFIPITA	IMFIPCAN
Correlation	0.899	0.889	0.767	0.831	0.940	0.946	0.731
CPI	IMFCPIUS	IMFCPIUK	IMFCPIJP	IMFCPIGER	IMFCPIFR	IMFCPIITA	IMFCPICAN
Correlation	0.805	0.810	0.761	0.525	0.887	0.891	0.739
PR	IMFFUNDUS	IMFFUNDUK	IMFFUNDJP	IMFFUNDCAN	IMFFUNDEU		
Correlation	0.844	0.911	0.330	0.914	0.073		

Notes: This table summarizes the correlations between the national macro series in Data Appendix and the global macro factors shown in Figure 8, for four categories: Financial Condition Index, Industrial Production growth rate, CPI and the change (YoY) of policy rate.

B.4.3 Co-Movement in Yields

B.4.3.1 Factor Dynamics

In this section, we depict the dynamics of the global yield factors estimated from our proposed 'Fundamentals-Augmented Hierarchical Factor Model'. As mentioned before, we extract two national yield factors that account for more than 96% of the variance of the term structure. We now focus on the global yield factors, as these factors typically drive the national Level and Slope factors. Firstly, we calculate the arithmetic sum of the global Level and Slope factors to evaluate the effect on the global short rate co-movement. This sum is denoted as the *global short rate factor*, and reflects the global co-movement in short rates across countries.[30] From the left panel of Figure 9, we can see the global short rate factor is strongly correlated with the first principal component of short rates across the seven advanced economies, also implying our model successfully captures the global co-movement of the short rates.[31] One feature of the movements of the global short rate factor is that it falls sharply after the Global Financial Crisis, consistent with a global expansion in monetary policy.

It is straightforward to decompose the global short rate factor into the global Level and Slope. The movements of these two factors are shown in the right panel of Figure 9, in which we also highlight some distinct historical events: January 1999 and the start of the euro area, US recessions in 2001 and 2008 as defined by NBER and the European sovereign debt crisis. As we have already discussed, Level and Slope factors control the shape of the term structure, which can be informative in revealing useful macroeconomic information. For example, before 1999 there is a downward trend for the Level factor and an upward trend for the Slope factor, which means the global term structure is moving down and flattening.[32] This phenomenon indicates a moderation in global term structure, possibly caused by greater integration.[33] We can observe two clear trends abstracting from temporary disturbances in the factors. Firstly, the downward-trending global Level seems to relate to the decreasing inflation level in the period of the Great Moderation, as suggested by Evans and Marshall (2007) and Koopman, Mallee and Van der Wel (2010). Secondly, the Slope factor is declining during US recessions, suggesting it is related to real economic activity, as indicated in Kurmann and Otrok (2013).

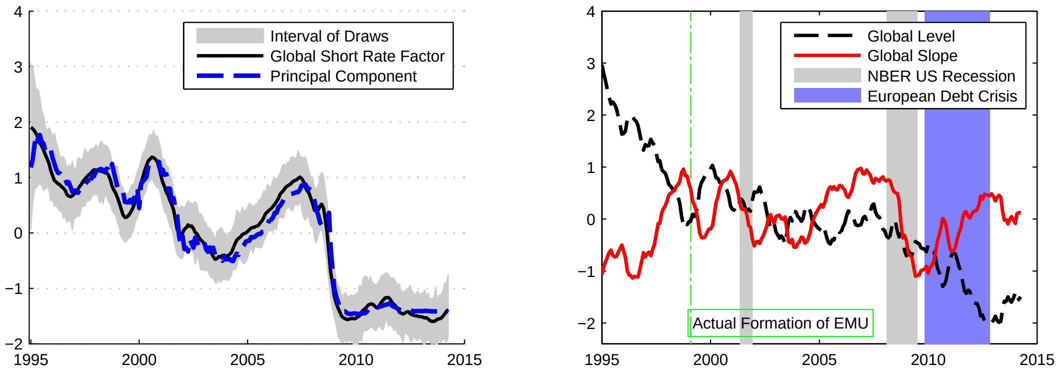
[30] By NS restrictions, for a bond at very short maturity, we have the equation that $short\ rate = \beta_1 L_t^{NS} + \beta_2 S_t^{NS}$, where the loadings equal to one, i.e. $\beta_1 = \beta_2 = 1$. Therefore, the short rate is directly driven by the sum of two factors in our model construction, see Appendix B.2.2 for details.

[31] Note that there is a smaller proportion of bond yield movements in country level that are not captured by the global yield factors. We find that these country-specific movements in national yield factors can be largely explained by the divergence of monetary policy in different countries. The results are consistent with the findings in Jotikasthira, Le and Lundblad (2015), but not shown here as we focus on the global co-movement.

[32] An increase in the level factor is consistent with higher yields on average. An increase in the slope factor is consistent with a flatter yield curve. In an extreme case, if two factor are moving in opposite directions but with the same magnitude, then the short rates stay still and long rates are driven by the changes in the Level factor.

[33] The strong negative correlation between the Level and Slop disappears after 1999 and reappears after the financial crisis.

Figure 9: *Global Short Rate Factor and the Decomposition*

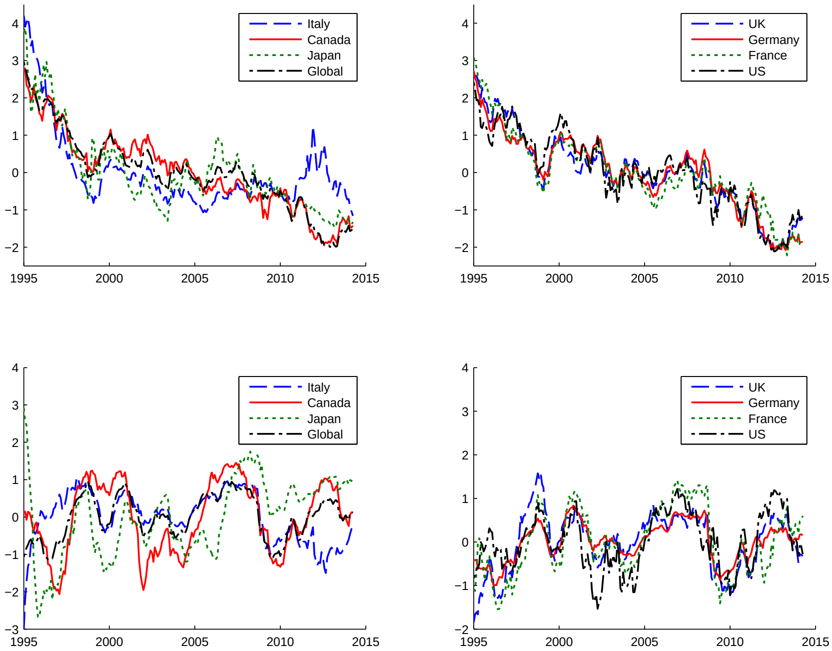


Notes: 1. The left panel shows the global short rate factor (i.e. an arithmetic sum of extracted global Level and Slope factors) and the first principal component of the national short-run policy rates (dashed line). The first principal component of national policy rates, which accounts for more than 84% of total variance of national policy rates. The gray areas cover all the draws of the global short rate factor (i.e. Level + Slope) from our model, and the solid black line is the median value of the draws. Data standardization implies yields can fall below zero.
2. The right panel shows the decomposition of the median of the global short rate factor. We decompose the short rate factor into the global Level (dashed line) and the global Slope (solid red line). In general, the Level factor controls the level of the term structure whereas the Slope factor controls the slope of the term structure. The shaded areas cover some major recession periods in the US and Europe.

B.4.3.2 Commonality of Level and Slope

We firstly plot our identified Level and Slope factors in Figure 10, respectively, in order to evaluate the commonalities in country-level yield factors. The Slope factors are relatively less persistent than the Level factors. From the figures it is evident that a strong co-movement in Level factor dynamics exists, but some also exists for the Slope. We also calculate the communality statistics for all countries in Table 16 to better quantify matters. That is we calculate the proportion of national level or slope factor explained by the global equivalent. This indicates that the commonality in Level factor dynamics is stronger but co-movement remains in the Slope. Generally, we find significant co-movement among Germany, France, Canada, UK and US. In contrast, the Level and Slope factors of Italy are relatively more divorced from the global factors, consistent with Table 3.2 above; the Japanese Slope factor is much less common among all Slope factors as the communality statistic is nearly zero. The above findings are reassuringly in line with the results in Diebold, Li and Yue (2008).

Figure 10: *Estimated Global and National Factors*



Notes: The upper panels show the median values of global Level and Slope factors and the national Level factors of Italy, Canada and Japan. The lower panels show the median values of the national Level and Slope factors of the UK, Germany, France and the US.

Table 16: *Communality Table of Level and Slope*

Level		Slope	
Country	Communality	Country	Communality
Italy	0.45	Italy	0.24
Canada	0.94	Canada	0.35
France	0.94	France	0.67
Germany	0.94	Germany	0.91
Japan	0.80	Japan	0.04
UK	0.98	UK	0.77
US	0.90	US	0.51
Average	0.85	**Average**	0.50

Notes: This table summarizes for all countries the communality statistics of global Level and Slope factors for national Level and Slope factors. For example, the communality for a given country is interpreted as the proportion of the variation in the national Level factor explained by the global Level factor. Likewise for the Slope communality.

B.4.4 Variance Decomposition across Maturities

Table 17: *Decomposition of Variance (US)*

Maturity (Month)	Posterior Mean (Standard Deviation)		
	$Share_G$	$Share_F$	$Share_X$
3	0.65(0.08)	0.32(0.08)	0.02(0.01)
6	0.68(0.08)	0.32(0.08)	0.01(0.00)
12	0.71(0.08)	0.29(0.08)	0.00(0.00)
24	0.74(0.07)	0.26(0.07)	0.01(0.00)
36	0.76(0.07)	0.24(0.07)	0.01(0.00)
48	0.77(0.07)	0.22(0.07)	0.01(0.00)
60	0.78(0.07)	0.22(0.06)	0.00(0.00)
72	0.79(0.06)	0.21(0.06)	0.00(0.00)
84	0.79(0.06)	0.21(0.06)	0.00(0.00)
96	0.79(0.06)	0.21(0.06)	0.01(0.00)
120	0.78(0.07)	0.20(0.06)	0.03(0.01)

Notes: This table summarizes the decomposition of variance for the three-level hierarchical model of US bond yields. $share_G$, $share_F$ and $share_Z$ denote the variance shares at different maturities in the country-level block of shocks ε_G, ε_F and ε_X, respectively. In each parenthesis $(\cdot)$ the posterior standard deviation of shares in a specific block is calculated from our draws, see Section 3.2. Larger standard deviation means higher uncertainty in the estimates, but we do not have an exact credible interval interpretation as the statistics do not necessarily follow (truncated) normal distributions.

Table 18: *Decomposition of Variance*

Maturity	UK			Germany			France		
(Month)	$Share_G$	$Share_F$	$Share_X$	$Share_G$	$Share_F$	$Share_X$	$Share_G$	$Share_F$	$Share_X$
3	0.80(0.06)	0.20(0.06)	0.01(0.00)	0.70(0.08)	0.23(0.06)	0.07(0.02)	0.66(0.08)	0.27(0.07)	0.07(0.02)
6	0.81(0.06)	0.19(0.06)	0.00(0.00)	0.71(0.08)	0.23(0.06)	0.06(0.02)	0.70(0.08)	0.28(0.07)	0.03(0.01)
12	0.83(0.05)	0.17(0.05)	0.01(0.00)	0.72(0.08)	0.23(0.07)	0.05(0.02)	0.73(0.07)	0.27(0.07)	0.00(0.00)
24	0.84(0.05)	0.14(0.05)	0.01(0.00)	0.75(0.08)	0.23(0.07)	0.02(0.01)	0.75(0.07)	0.24(0.07)	0.01(0.00)
36	0.86(0.05)	0.13(0.04)	0.01(0.00)	0.76(0.07)	0.23(0.07)	0.01(0.00)	0.77(0.07)	0.22(0.06)	0.02(0.00)
48	0.88(0.04)	0.12(0.04)	0.00(0.00)	0.77(0.07)	0.23(0.07)	0.00(0.00)	0.78(0.06)	0.21(0.06)	0.01(0.00)
60	0.89(0.04)	0.11(0.04)	0.00(0.00)	0.77(0.07)	0.22(0.07)	0.01(0.00)	0.79(0.06)	0.20(0.06)	0.00(0.00)
72	0.89(0.04)	0.11(0.04)	0.00(0.00)	0.76(0.07)	0.21(0.07)	0.02(0.01)	0.80(0.06)	0.20(0.06)	0.00(0.00)
84	0.88(0.04)	0.10(0.04)	0.01(0.00)	0.75(0.08)	0.21(0.06)	0.04(0.01)	0.80(0.06)	0.20(0.06)	0.00(0.00)
96	0.86(0.04)	0.10(0.03)	0.04(0.01)	0.73(0.08)	0.20(0.06)	0.06(0.02)	0.80(0.06)	0.19(0.06)	0.01(0.00)
120	0.80(0.06)	0.09(0.03)	0.11(0.03)	0.71(0.08)	0.19(0.06)	0.10(0.03)	0.78(0.06)	0.18(0.05)	0.04(0.01)

Notes: This table summarizes the decomposition of variance for the three-level hierarchical model of bond yields. For each country, $share_G$, $share_F$ and $share_Z$ denote the variance shares at different maturities in the country-level block of shocks ε_G, ε_F and ε_X, respectively. In each parenthesis $(\cdot)$ the posterior standard deviation of shares in a specific block is calculated.

Table 19: *Decomposition of Variance (Continued)*

Maturity	Italy			Canada			Japan		
(Month)	$Share_G$	$Share_F$	$Share_X$	$Share_G$	$Share_F$	$Share_X$	$Share_G$	$Share_F$	$Share_X$
3	0.31(0.09)	0.66(0.09)	0.03(0.01)	0.52(0.10)	0.36(0.08)	0.12(0.03)	0.5(0.10)	0.44(0.09)	0.06(0.01)
6	0.32(0.10)	0.67(0.09)	0.01(0.00)	0.57(0.09)	0.36(0.08)	0.07(0.02)	0.54(0.10)	0.43(0.09)	0.03(0.01)
12	0.34(0.10)	0.66(0.10)	0.00(0.00)	0.63(0.09)	0.35(0.08)	0.02(0.01)	0.60(0.09)	0.39(0.09)	0.02(0.00)
24	0.35(0.10)	0.64(0.10)	0.00(0.00)	0.70(0.08)	0.30(0.08)	0.00(0.00)	0.65(0.08)	0.31(0.08)	0.04(0.01)
36	0.36(0.10)	0.63(0.10)	0.00(0.00)	0.74(0.07)	0.26(0.07)	0.00(0.00)	0.69(0.08)	0.28(0.07)	0.03(0.01)
48	0.37(0.10)	0.62(0.10)	0.00(0.00)	0.76(0.07)	0.24(0.07)	0.00(0.00)	0.72(0.07)	0.26(0.07)	0.02(0.00)
60	0.38(0.10)	0.62(0.10)	0.00(0.00)	0.77(0.07)	0.23(0.07)	0.00(0.00)	0.75(0.07)	0.25(0.07)	0.01(0.00)
72	0.38(0.10)	0.61(0.10)	0.00(0.00)	0.78(0.06)	0.22(0.06)	0.00(0.00)	0.76(0.07)	0.24(0.07)	0.00(0.00)
84	0.39(0.10)	0.61(0.10)	0.01(0.00)	0.79(0.06)	0.21(0.06)	0.00(0.00)	0.76(0.07)	0.23(0.06)	0.01(0.00)
96	0.39(0.10)	0.60(0.10)	0.01(0.00)	0.79(0.06)	0.21(0.06)	0.00(0.00)	0.75(0.07)	0.22(0.06)	0.02(0.01)
120	0.39(0.10)	0.59(0.10)	0.02(0.00)	0.79(0.06)	0.20(0.06)	0.01(0.00)	0.73(0.07)	0.21(0.06)	0.06(0.02)

Notes: This table summarizes the decomposition of variance for the three-level hierarchical model of bond yields. For each country, $share_G$, $share_F$ and $share_Z$ denote the variance shares at different maturities in the country-level block of shocks ε_G, ε_F and ε_X, respectively. In each parenthesis $(\cdot)$ the posterior standard deviation of shares in a specific block is calculated.

B.5 Robustness of Spillover Effects

Table 20 displays the correlation matrix of the country-specific components in national Level and Slope factors, which implies potential 'Granger causality' among country-specific components.

We set strict criteria for the 'Granger causality' to reveal 'spillovers'. The critical value of the test is set to be 0.01, and the maximum lag is set to be one as the transmission in financial market is considered very rapid. We conduct the causality test for all the draws obtained from our model. We then construct two directed graphs according to the results of Granger causality test in Figure 11.[34]

The upper graphs in Figure 11 display the asymmetric 'spillovers' among Level factors. One obvious observation is that the country-specific components of the UK Granger-cause the country-specific movements in Level factors of all other countries, which implies that the country-specific movements of the UK bond factors release some signals to other markets and cause different degrees of shifts in term structures. But the interpretations of the signals are heterogeneous in different markets, so the 'spillovers' are not captured by the global co-movement.

The lower graphs in Figure 11, in contrast, display the asymmetric 'spillovers' among Slope factors. It is evident that the country-specific movements of Italy in Slope factor are susceptible to all the country-specific components of other countries, which suggests the vulnerability of the Italy bond market.[35]

More interesting observations are shown in Figure 11. Regarding the sovereign risks of Italy, it seems that the risks can influence the levels of bond yields of the US and Germany, but the contagion to Germany market is more evident, as it also affects the movements in Slope. It is possible that the spillovers from Italy to the US arise through the Germany market. Regarding the bond market of Japan, it is clear that the market is closely connected to the US market, as the movements in the US Granger-cause the movements of Japan in both Level and Slope.[36] Nevertheless, the response from Japan is not significant as the US market is solely affected by European markets in Level.

In terms of the reflective mechanism among the markets, we should pay attention to three pairs: UK-France, Japan-Canada and Italy-Germany. However, to confirm whether there are amplifications or counteractions, more work needs to be done. Moreover, a more complicated mechanism lies in the relation between the movements in the UK, Italy and Canada and the reactions in Germany; the first three countries are likely to affect the Level of Germany, while the feedback from Germany is through the effects on the Slope of the three.

[34] The test results can be found in Table 21 and 22.

[35] Generally, the changes in Slope factor, i.e. the changes in the shape of the term structure, potentially reflect more severe changes in investor sentiment than the case of parallel shifts.

[36] The reason why the US country-specific movements in yield factors do not affect other markets is that the 'fundamental' effects from the US have been captured in the global factor movements.

Table 20: *Correlation Matrix of the Country-Specific Components*

	ITA_L	ITA_S	CAN_L	CAN_S	FRA_L	FRA_S	GER_L	GER_S	JP_L	JP_S	UK_L	UK_S	US_L
ITA_L													
ITA_S	-0.65(0.03)												
CAN_L													
CAN_S			-0.47(0.04)										
FRA_L	0.49(0.07)												
FRA_S													
GER_L	-0.34(0.07)				0.46(0.05)	0.40(0.07)							
GER_S													
JP_L	0.47(0.04)		-0.41(0.06)	0.37(0.03)			-0.42(0.04)						
JP_S		-0.47(0.03)					0.31(0.05)						
UK_L													
UK_S	-0.34(0.07)												
US_L	-0.57(0.04)				-0.31(0.07)								
US_S	0.47(0.04)		-0.44(0.05)	0.55(0.03)			-0.36(0.06)		0.65(0.03)				-0.34(0.05)

Notes: 1. This table summarizes the correlation matrix of the country-specific components among the Level and Slope factors of all countries: Italy (ITA), Canada (CAN), France (FRA), Germany (GER), Japan (JP), the UK and the US. In each parenthesis $(\cdot)$ the posterior standard deviation of the correlation element is calculated from our draws, see Section 3.2. Larger standard deviation means higher uncertainty in the estimates, but we do not have an exact credible interval interpretation as the statistics do not necessarily follow (truncated) normal distributions.
2. The diagonal elements are dismissed and we only show the elements in the lower triangular part of the correlation matrix with absolute values larger than 0.30. Subscripts *L* and *S* are for Level and Slope factors respectively.

Table 21: *Granger Causality of the Country-Specific Components (Level)*

	ITA_L	CAN_L	FRA_L	GER_L	JP_L	UK_L	US_L
ITA_L	NA			***			***
CAN_L		NA	**	**	***		
FRA_L			NA			**	***
GER_L				NA	***		*
JP_L		**			NA		
UK_L	*	**	***	***	*	NA	***
US_L					**		NA
ITA_S							
CAN_S							
FRA_S							
GER_S							
JP_S							
UK_S							***
US_S					***		

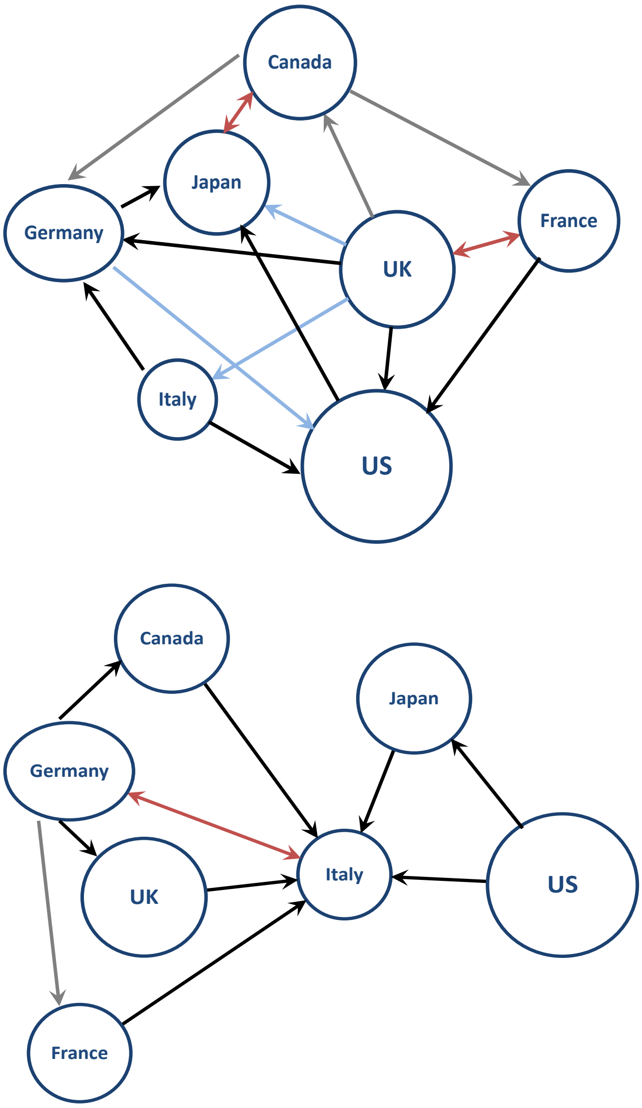
Notes: 1. This table summarizes the Granger causality statistics of the Country-Specific Components of all countries: Italy (ITA), Canada (CAN), France (FRA), Germany (GER), Japan (JP), the UK and the US. The subscript of each country indicates the factor Level (L) or Slope (S).
2. The diagonal elements in the upper half table are not applicable. Each column indicates whether the component of one country is Granger-caused by other components.
3. The significance level of the Granger causality test is set to be 0.01, and the lag is set to be 1. *, ** and *** indicate 70%, 80% and 90% of the posterior draws reject the test, respectively.

Table 22: *Granger Causality of the Country-Specific Components (Slope)*

	ITA_S	CAN_S	FRA_S	GER_S	JP_S	UK_S	US_S
ITA_L	***				***		***
CAN_L	***						
FRA_L	***						
GER_L	***		**				
JP_L	***						
UK_L	***					**	
US_L	***				***		
ITA_S	NA						
CAN_S	***	NA					
FRA_S	***		NA				
GER_S	***	***		NA		***	
JP_S	***				NA		
UK_S	***					NA	
US_S	***						NA

Notes: 1. This table summarizes the Granger causality statistics of the Country-Specific Components of all countries: Italy (ITA), Canada (CAN), France (FRA), Germany (GER), Japan (JP), the UK and the US. The subscript of each country indicates the factor Level (L) or Slope (S).
2. The diagonal elements in the lower half table are not applicable. Each column indicates whether the component of one country is Granger-caused by other components.
3. The significance level of the Granger causality test is set to be 0.01, and the lag is set to be 1. *, ** and *** indicate 70%, 80% and 90% of the posterior draws reject the test, respectively.

Figure 11: *Directed Graphs of 'Spillovers' in Country-Specific Components*



Notes: The upper figure shows how each country-specific component in Level are affected by components of other countries, whereas the lower figure displays the influence in Slope. The graphs are constructed according to the results in Table 21 and 22.

CHAPTER 4

Learning about Term Structure Predictability under Uncertainty

ABSTRACT

This paper proposes a no-arbitrage framework of term structure modeling with learning and model uncertainty. The representative agent considers parameter instability, as well as the uncertainty in learning speed and model restrictions. The empirical evidence shows that apart from observational variance, parameter instability is the dominant source of predictive variance when compared with uncertainty in learning speed or model restrictions. When accounting for ambiguity aversion, the out-of-sample predictability of excess returns implied by the learning model can be translated into significant and consistent economic gains over the Expectations Hypothesis benchmark.

Keywords: Affine Term Structure Models, Learning, Parameter Uncertainty, Model Uncertainty, Ambiguity Aversion, Bayesian Methods.
JEL Classification Codes: C1, C3, C5, D8, E4, G1.

Author Contributions: This is my job market paper.

4.1 Introduction

Modeling the interest rate term structure is essential in understanding expectations of risk compensation and the future path of monetary policy. For instance, the affine class of arbitrage-free term structure models has gained great popularity in both pricing and predicting future movements of bonds, because of its parsimonious factor structure and tractability. As a stylized fact, the predictability of bond returns is widely recognized in Fama and Bliss (1987), Cochrane and Piazzesi (2005), Sarno, Thornton and Valente (2007) and Ludvigson and Ng (2009). The traditional Expectations Hypothesis (EH) has been strongly rejected statistically and therefore, the term premia should be time-varying.[1] Accurate term premium predictions should be useful for portfolio optimization, as it guides a mean-variance investor to making a tradeoff between expected returns and the volatility of the portfolio.

Surprisingly, significant predictability in expected bond returns cannot be translated into large economic gains, as suggested by Della Corte, Sarno and Thornton (2008), Thornton and Valente (2012) and Sarno, Schneider and Wagner (2014). They reach the conclusion that when compared with a EH investor, the investor using alternative prediction models with statistical significance cannot improve economic utility. The seemingly contradictory evidence in this literature is indeed puzzling. Seeking to resolve this puzzle with those findings, Gargano, Pettenuzzo and Timmermann (2014) allow for parameter and model uncertainty. However, the resolution to the puzzle is far from perfect and further research is required. To resolve the *economic value puzzle*, it is necessary to understand the uncertainty in the predictability and, moreover, to consider various sources of uncertainty when making the optimal portfolio choice.

This issue is revisited by taking account of both parameter and model uncertainty. We propose a flexible term structure model which includes time-varying coefficients, stochastic volatility and dynamic model selection. These features are sensible in an agent's pricing and forecasting problem because parameters and models are uncertain without sufficient data. Like an econometrician, the agent needs to learn about the evolution of the state of the economy (Cagetti et al. (2002) and Hansen (2007)). These features can be formalized in an affine model to accommodate structural changes, where the learning speed or 'gain' is specified.[2]

The seminal contribution of Timmermann (1993, 1996) studies the implications of learning in explaining the volatility and predictability of asset returns. There is substantial literature employing learning to explain a range of financial market anomalies. Specifically, Piazzesi and Schneider (2007) and Collin-Dufresne, Johannes and Lochstoer (2013) examine the implications of learning in a preference-based asset pricing framework, and they show that learning can explain standard puzzles in bond yields. Using a reduced-form pricing kernel, Kozicki and Tinsley (2001) and Dewachter and

[1] A weak form of EH requires the term premia to be a constant, which implies that expected excess bond returns should not be predictable.

[2] As shown in Evans and Honkapohja (2001), this framework provides an expectationally stable solution as long as the gain parameter is sufficiently small.

Lyrio (2008) study the learning problem in which agents continuously update their beliefs regarding the central bank's policy targets, but they only allow for time variations of the drift parameter. Laubach, Tetlow and Williams (2007), Orphanides and Wei (2012) and Cieslak and Povala (2014) relax the assumptions about the potential sources of structural instability and allow for updating beliefs of all model parameters. A common practice is to use macro variables as pricing factors, which may cause undesired mispricing as indicated by Anh and Joslin (2013). To avoid the potential mispricing problem and increase predictive power, in this paper we consider portfolios as risk factors, similar to Joslin, Singleton and Zhu (2011).

However, learning does not guarantee the convergence of agents' heterogeneous beliefs. Bond yields are highly persistent, and Kurz (1994) suggests that if the economic system is close to nonstationary, limited data would make it difficult for rational investors to identify the correct model from alternative ones. Model uncertainty can arise from the imposition of restrictions related to model identification. In order to increase forecast performance, researchers impose over-identifying restrictions motivated statistically or economically.[3] The economic dynamics are ambiguous with undetermined restrictions and therefore, statistical methods are employed to determine the optimal specification. Joslin, Priebsch and Singleton (2014) and Jotikasthira, Le and Lundblad (2015) choose restrictions based on the Bayesian Information Criterion (BIC), while Bauer (2015) uses Bayesian model averaging method to calculate the weighted average across specifications for robust inference. The above model selections are conducted with full-sample data, and a real-time dynamic model selection is desirable in the topic of interest rate forecasting. In stock return predictions, Cremers (2002) and Avramov (2002) have shown that allowing investors to dynamically select between different models is useful to control for the data snooping problem and can increase out-of-sample predictability. More recent studies, for instance, Dangl and Halling (2012), Johannes, Korteweg and Polson (2013) and Gargano, Pettenuzzo and Timmermann (2014) adopt Bayesian approaches to accommodate model uncertainty in stock and bond return forecasts. In this paper, we adopt the same framework as Dangl and Halling (2012) to conduct a real-time Bayesian model selection, explicitly exploring model uncertainty in term structure modeling.

Uncertainty in parameters and models is not explicitly priced in classical term structure models, but a sophisticated investor should be aware of the uncertainty when making investment decisions. Pástor and Stambaugh (1999, 2000) investigate how the uncertainty in parameters or models changes the way we make portfolio decisions.[4] Investors prefer known risks over unknown risks, so Uppal and Wang (2003) introduce an important extension to allow for ambiguity aversion. While most of the related research focuses on the portfolio allocation of stocks, very few recent papers have approached the topic of bond returns.[5] In order to close this gap, we consider a generalized framework with

[3] See for example, Christensen, Diebold and Rudebusch (2011), Duffee and Stanton (2012) and Joslin, Priebsch and Singleton (2014).

[4] Recent contributions on portfolio choice under uncertainty include Brandt et al. (2005), Avramov and Chordia (2006), and Rapach, Strauss and Zhou (2009).

[5] Gagliardini, Porchia and Trojani (2009) and Ulrich (2013) show that considering ambiguity aversion helps explain the term premia dynamics but they do not explicitly explore the portfolio allocation problem. Della Corte, Sarno and Thornton (2008), Thornton and Valente (2012) and Sarno, Schneider and Wagner (2014) study the bond portfolio selection problem for risk-averse investors. Gargano, Pettenuzzo and Timmermann

ambiguity aversion that nests the case of ordinary risk-averse investors, following and extending Garlappi, Uppal and Wang (2007).

This paper builds upon the work of Giacoletti, Laursen and Singleton (2014), who construct a learning framework of arbitrage-free affine term structure models and who investigate different learning rules in term structure forecasts. We further introduce model uncertainty in the learning problem, which provides flexibility in selecting the best restrictions imposed on factor dynamics and the optimal learning gain/speed. More importantly, this extension allows the analysis of the uncertainty in the predictive performance in order to reveal the sources of prediction uncertainty. Our work is also related to Gargano, Pettenuzzo and Timmermann (2014), who evaluate the economic gains of models with parameter and model uncertainty, but differs in a way that we consider a more generalized portfolio allocation problem with ambiguity aversion. In this framework, we explore to what extent investors benefit from ambiguity aversion in addition to the traditional risk aversion.

In particular, the proposed learning model nests most of the affine term structure models with learning, and is flexible in selecting the optimal specification from different learning speeds and model restrictions. Utilizing our approach we make several contributions to understanding the US bond market from 1961:06 to 2014:10. The first finding is that the pricing dynamics have not varied much since the 1960s, which is consistent with Giacoletti, Laursen and Singleton (2014), but we observe large variability in factor dynamics under the physical measure. The proposed model is promising in forecasting, as its predictive performance using conditional information is similar to the benchmark model using full information. By analyzing the sources of predictive uncertainty, it can be seen that, apart from observational variance, parameter instability is the main driver of predictive variance. Uncertainty in learning speed or model specification, *vis-à-vis* parameter instability, does not generally play an important role.

With respect to asset allocation, we consider both parameter and model uncertainty by extending the mean-variance framework proposed by Garlappi, Uppal and Wang (2007). Our ambiguity-averse investor successfully turns the predictability of excess returns implied by the learning model into substantial economic gains, when compared with the Expectations Hypothesis benchmark. In addition to parameter uncertainty, the consideration of model uncertainty is the key to ensuring success. This finding is robust compared to different subsample periods, despite that the economic gains can be eroded during the financial crisis. Therefore, this framework resolves the *economic value puzzle* in bond return predictions with the evidence in the previous term structure literature.

The rest of the paper is structured as follows. Section 4.2 describes the methodology, the term structure models considered, and the framework with ambiguity aversion for evaluating the out-of-sample predictability of excess returns. Section 4.3 outlines the empirical results of the learning model and its out-of-sample portfolio performance, including discussion about pricing dynamics, physical dynamics and term structure predictability. Section 4.4 concludes.

(2014) evaluate the economic gains by considering parameter and model uncertainty, but also for risk-averse investors.

4.2 Methodology

4.2.1 A Canonical Gaussian Dynamic Term Structure Model (GDTSM)

We firstly consider an economic environment in which agents value nominal bonds using the stochastic discount factor or pricing kernel. The one-period pricing kernel or stochastic discount factor of an asset is given by

$$\mathcal{M}_{Z,t+1} = e^{-r_t - \frac{1}{2}\Lambda'_{Zt}\Lambda_{Zt} - \Lambda'_{Zt}\varepsilon^{\mathbb{P}}_{t+1}}, \tag{4.1}$$

where the $N \times 1$ state vector Z_t encompasses all risks in the economy, Λ_{Zt} is the vector collecting market prices of risk, and r_t is the one-period bond yield. In the absence of macro risks, Z_t is a linear rotation of $N \times 1$ vector of portfolio risk factors $\mathcal{P}_t$.[6]

Following Joslin, Singleton and Zhu (2011), we specify the pricing kernel in the bond market. The bond-market-specific $\mathcal{M}_{P,t+1}$, conditional on the information of the priced risks in the bond market $\mathcal{P}_t$,[7] is now given by

$$\mathcal{M}_{\mathcal{P},t+1} = e^{-r_t - \frac{1}{2}\Lambda'_{\mathcal{P}t}\Lambda_{\mathcal{P}t} - \Lambda'_{\mathcal{P}t}\varepsilon^{\mathbb{P}}_{\mathcal{P},t+1}}, \tag{4.2}$$

where the short rate r_t is an affine function of $\mathcal{P}_t$,

$$r_t = \rho_{0\mathcal{P}} + \rho_{1\mathcal{P}} \cdot \mathcal{P}_t, \tag{4.3}$$

and the risks $\varepsilon^{\mathbb{P}}_{\mathcal{P},t+1}$ are the N innovations from the unconstrained first-order vector-autoregressive (VAR) model under the physical or historical measure $\mathbb{P}$[8]

$$\mathcal{P}_t = K^{\mathbb{P}}_{0\mathcal{P}} + K^{\mathbb{P}}_{\mathcal{P}\mathcal{P}}\mathcal{P}_{t-1} + \sqrt{\Sigma_{\mathcal{P}\mathcal{P}}}\varepsilon^{\mathbb{P}}_{\mathcal{P}t}, \tag{4.4}$$

where $\varepsilon^{\mathbb{P}}_{\mathcal{P}t} \sim N(0, I_N)$ and $\Sigma_{\mathcal{P}\mathcal{P}}$ is an $N \times N$ nonsingular matrix. We close the model by specifying the dynamics of $\mathcal{P}_t$ under the pricing (risk-neutral) distribution $\mathbb{Q}_t$

$$\mathcal{P}_t = K^{\mathbb{Q}}_{0\mathcal{P}} + K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}\mathcal{P}_{t-1} + \sqrt{\Sigma_{\mathcal{P}\mathcal{P}}}\varepsilon^{\mathbb{Q}}_{\mathcal{P}t}. \tag{4.5}$$

Under the above assumptions and the absence of arbitrage opportunities, the yield on an m-

[6]Our model can be easily extended to a setup with unspanned macro risks where Z_t includes the information of macroeconomic risk factors M_t in addition to portfolio risk factors $\mathcal{P}_t$.

[7]Without loss of generality, we rotate the N risk factors to make normalization. Accordingly, $\mathcal{P}_t$ corresponds to the N portfolios of bond yields; for example, $\mathcal{P}$ can be the first N principal components (PCs) of bond yields. Joslin, Singleton and Zhu (2011) show that the rotation is normalized so that the parameters governing the $\mathbb{Q}$ distribution of yields, i.e. $(\rho_{0\mathcal{P}}, \rho_{1\mathcal{P}}, K^{\mathbb{Q}}_{0\mathcal{P}}, K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}})$ are fully determined by the parameter set $(\Sigma_{\mathcal{P}\mathcal{P}}, \lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_{\infty})$, where $\lambda^{\mathbb{Q}}$ denotes the N-vector of ordered nonzero eigenvalues of $K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}$ and $r^{\mathbb{Q}}_{\infty}$ denotes the long-run mean of r_t under $\mathbb{Q}$.

[8]This representation, which characterizes the dynamics of the full set of N risk factor, can be viewed as the companion form of a higher-order VAR of the state vector of risk factors.

period bond, for any $m > 0$, is an affine function of $\mathcal{P}_t$,

$$y_t^m = A_{\mathcal{P}}(m) + B_{\mathcal{P}}(m)\mathcal{P}_t, \tag{4.6}$$

where the loadings $A_{\mathcal{P}}(m)$ and $B_{\mathcal{P}}(m)$ govern the $\mathbb{Q}$ distribution of yields.[9] The detailed expressions of the loadings can be found in Appendix C.1.

The scaled market prices of risk are also affine functions of $\mathcal{P}_t$,

$$\Sigma_{\mathcal{P}\mathcal{P}}^{1/2}\Lambda_{\mathcal{P}}(\mathcal{P}_t) = \Lambda_0 + \Lambda_1\mathcal{P}_t, \tag{4.7}$$

where $\Lambda_1 = K_{\mathcal{P}\mathcal{P}}^{\mathbb{P}} - K_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}}$ is an $N \times N$ matrix and $\Lambda_0 = K_{0\mathcal{P}}^{\mathbb{P}} - K_{0\mathcal{P}}^{\mathbb{Q}}$ is an $N \times 1$ vector.

4.2.2 Learning and Model Uncertainty

From the last section, we can see that the bond-market-specific pricing kernel $\mathcal{M}_{\mathcal{P},t+1}$ is a function of priced risks $\mathcal{P}$ and a set of parameters $\Theta \equiv (\Theta^{\mathbb{P}}, \Theta^{\mathbb{Q}})$ that govern the dynamics under the physical measure $\mathbb{P}$ and risk-neutral measure $\mathbb{Q}$. To be more specific, the parameter set $\Theta^{\mathbb{P}} \equiv (K_{0\mathcal{P}}^{\mathbb{P}}, K_{\mathcal{P}\mathcal{P}}^{\mathbb{P}})$ governs the drift of $\mathcal{P}$ under the physical measure, whereas the set $\Theta^{\mathbb{Q}} \equiv (\Sigma_{\mathcal{P}\mathcal{P}}, \lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$ determines the risk-neutral dynamics, i.e. the pricing distribution. Note that the variance matrix $\Sigma_{\mathcal{P}\mathcal{P}}$ in fact enters both the physical and risk-neutral dynamics, which can be estimated from Equation (4.4) that describes the physical dynamics of pricing factor $\mathcal{P}$.

We consider a representative agent who can adaptively learn about the evolution of the state of the economy. He or she may have different perceptions of the pricing kernel $\mathcal{M}_{\mathcal{P},t+1}$ at different points in time.[10] As mentioned in Evans and Honkapohja (2001), the concept of *adaptive learning* (AL) introduces a specific form of bounded rationality, and provides a means of approximating agents' expectations that incorporates learning as well as a rationale for rational expectations. Based on the learning concept, we rewrite the evolution of the one-period pricing kernel under the physical measure $\mathbb{P}$, conditional on the information at time t, as

$$\mathcal{M}_{t,t+1} = E_t^{\mathbb{P}}[f_{\mathcal{M}}(\Theta_t, \mathcal{P}_{t+1})|\mathcal{P}_t] = F_{\mathcal{M}}(\Theta_t, \mathcal{P}_t). \tag{4.8}$$

Therefore, the price D_t^m of a zero-coupon bond issued at date t and maturing at date m, is also a function of $(\Theta_t, \mathcal{P}_t)$ under the physical measure $\mathbb{P}$

$$D_t^m = E_t^{\mathbb{P}}[\prod_{s=1}^{m} \mathcal{M}_{t+s-1,t+s}] = F_{D_m}(\Theta_t, \mathcal{P}_t). \tag{4.9}$$

[9]As we will see in the next section, the loadings are known functions of parameters $\Theta^{\mathbb{Q}} \equiv (\Sigma_{\mathcal{P}\mathcal{P}}, \lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$.

[10]Similar to Giacoletti, Laursen and Singleton (2014), we consider a model with a reduced-form pricing kernel, which does not clearly specify agents' preferences when compared with preference-based models such as Piazzesi and Schneider (2007) and Collin-Dufresne, Johannes and Lochstoer (2013). Nevertheless, as shown in Duffie (2001) and Piazzesi (2010), we can link the pricing equation to fundamentals within a representative agent endowment economy where preference parameters are specified.

To simplify the estimation problem in our learning system, we have the following assumptions:

- The portfolio risk factors $\mathcal{P}_t$ are measured without errors.
- The parameters Θ, which may evolve over time, are unknown to the agent, and hence, need to be estimated statistically at each point in time t.
- The risk of unknown parameters is not priced.

These assumptions are standard in the literature of term structure pricing or learning, see Joslin, Singleton and Zhu (2011), Joslin, Priebsch and Singleton (2014) and Giacoletti, Laursen and Singleton (2014). With these mild assumptions, we can partition the parameter set Θ_t into subsets $\Theta^{\mathbb{P}}$ and $\Theta^{\mathbb{Q}}$ and estimate them respectively.

For the physical dynamics, we consider a case where the agent believes the law of motion (*perceived law of motion*) of parameters $\Theta^{\mathbb{P}}$ is a random walk process, and then Equation (4.4) becomes

$$\mathcal{P}_t = K^{\mathbb{P}}_{t,0\mathcal{P}} + K^{\mathbb{P}}_{t,\mathcal{P}\mathcal{P}}\mathcal{P}_{t-1} + \sqrt{\Sigma_{t,\mathcal{P}\mathcal{P}}}\varepsilon^{\mathbb{P}}_{\mathcal{P}t}, \tag{4.10}$$

$$\begin{bmatrix} K^{\mathbb{P}}_{t,0\mathcal{P}} \\ \mathrm{vec}(K^{\mathbb{P}}_{t,\mathcal{P}\mathcal{P}}) \end{bmatrix} = \begin{bmatrix} K^{\mathbb{P}}_{t-1,0\mathcal{P}} \\ \mathrm{vec}(K^{\mathbb{P}}_{t-1,\mathcal{P}\mathcal{P}}) \end{bmatrix} + u_t, \tag{4.11}$$

where $\mathrm{vec}(\cdot)$ means the vectorization of a matrix and u_t is a vector of transition errors. The above system can be estimated using a (Bayesian) Kalman filter.

For the pricing dynamics, the *perceived law of motion* of parameters $(r^{\mathbb{Q}}_{\infty}, \lambda^{\mathbb{Q}})$ in $\Theta^{\mathbb{Q}}$ is also a random walk process. We rewrite Equation (4.6) as

$$y^m_t = A^m_{\mathcal{P}}(\Sigma_{t,\mathcal{P}\mathcal{P}}, \lambda^{\mathbb{Q}}_t, r^{\mathbb{Q}}_{t,\infty}) + B^m_{\mathcal{P}}(\lambda^{\mathbb{Q}}_t)\mathcal{P}_t, \tag{4.12}$$

$$\begin{bmatrix} r^{\mathbb{Q}}_{t,\infty} \\ \lambda^{\mathbb{Q}}_t \end{bmatrix} = \begin{bmatrix} r^{\mathbb{Q}}_{t-1,\infty} \\ \lambda^{\mathbb{Q}}_{t-1} \end{bmatrix} + \mathrm{u}_t, \tag{4.13}$$

where $\Sigma_{t,\mathcal{P}\mathcal{P}}$ is estimated from Equation (4.10) and u_t is a vector of transition errors, see Joslin, Singleton and Zhu (2011) for technical details. We estimate the above nonlinear system with the unscented Kalman filter.[11]

4.2.2.1 Learning Rules

Let us start with the physical dynamics. For a more convenient description, we rewrite the learning dynamics (4.10) and (4.11) under the physical measure as a form of p-lag time-varying parameter

[11]For identification one can fix $r^{\mathbb{Q}}_{\infty} = 0$, see for example Dai and Singleton (2000), Christensen, Diebold and Rudebusch (2011) and Joslin, Singleton and Zhu (2011).

vector autoregression (TVP-VAR)[12]

$$z_t = X_t \beta_t + v_t^{\mathbb{P}}, \tag{4.14}$$

$$\beta_{t+1} = \beta_t + u_t, \tag{4.15}$$

where $z_t = \mathcal{P}_t$, $X_t = I_N \otimes \left[z'_{t-1}, ..., z'_{t-p}\right]$, $\beta_t = \left[c_t, vec\left(B_{1t}\right)', \cdots, vec\left(B_{pt}\right)'\right]'$ is a vector summarizing all VAR coefficients, $v_t^{\mathbb{P}} \sim N\left(0, \Sigma_t\right)$ with Σ_t an $n \times n$ measurement covariance matrix, and $u_t \sim N\left(0, Q_t\right)$ with an $n \times n$ transition covariance matrix.

As we have mentioned, the system can be solved by means of the Kalman filter, see Appendix C.2.1 for details. The solution for this system follows a recursive rule given by

$$\beta_t | D_t \sim N\left(m_t, \Phi_t\right), \tag{4.16}$$

where D_t is the information set at time t. The solution is equivalent to a special case of the class of *adaptive least squares* (ALS) learning proposed by McCulloch (2007), which also nests the *ordinary least squares* (OLS) and *constant gain least squares* (CGLS) algorithm introduced by Sargent (2002) and Evans and Honkapohja (2001). The ALS formulas are given by

$$m_t = m_{t|t-1} + R_t^{-1} X_t' \Sigma_t^{-1} \widetilde{v}_t, \tag{4.17}$$

$$R_t = (1 - \gamma_t) R_{t-1} + X_t' \Sigma_t^{-1} X_t, \tag{4.18}$$

where $\widetilde{v}_t = z_t - X_t m_{t|t-1}$ is the prediction error and γ_t is the gain parameter which belongs to interval $[0, 1)$. Note that apart from the learning gain, stochastic volatility also plays a role in controlling the informativeness of incoming information flows, which parallels the finding of Cieslak and Povala (2015*b*) that stochastic volatility has a non-trivial effect on the conditional distribution of interest rates.

By setting the gain parameter to different values, we have different learning algorithms or rules:

- **Learning Rule 1**: When $\gamma_t = Q_t\left(\Phi_{t-1} + Q_t\right)^{-1}$ and $R_t = \Phi_t^{-1}$, the learning algorithm is the most general case of the ALS, i.e. the standard Kalman filter solution.[13]
- **Learning Rule 2**: When γ_t is replaced by a sufficiently small constant, as in Sargent (2002) and Evans and Honkapohja (2001), the learning rule becomes the *constant gain least squares* (CGLS) algorithm. This case is also consistent with the ‘forgetting factor’ algorithm proposed by Koop and Korobilis (2012, 2013), see Appendix C.2.1.
- **Learning Rule 3**: When $\gamma_t = 0$, the learning algorithm becomes the *recursive least squares* (RLS), i.e. a recursive form of *ordinary least squares* (OLS).

We will focus on the last two learning rules. In learning rule 3, with $\gamma = 0$ we immediately get

[12]Note that p is usually set to 1 in most of the no-arbitrage affine term structure models.
[13]The derivation of this result is provided in Appendix C.2.2

$Q_t = 0$ (Appendix C.2.2), so the ALS degenerates to a constant parameter case, or a *decreasing gain* case in Evans and Honkapohja (2001). When there are no structural changes, m_t will converge to the true value when $t \to \infty$. However, when compared with the constant gain case, the decreasing gain case has a lower convergence speed. Moreover, gain sequences decrease to zero in the constant gain case when $t \to \infty$, so this model cannot sufficiently deal with structural changes. Therefore, we need to consider learning rule 2 with constant gains and face a trade-off: A larger constant gain is better at tracking changes but at the cost of larger variance. Hence, similar to Sargent (2002), we only consider small gains to avoid instability.[14]

For the nonlinear system of the pricing dynamics, we can still write the rules of *adaptive learning* similar to the formulas of the Kalman filter, see Appendix C.2.3.

4.2.2.2 Model Uncertainty and Dynamic Model Selection

On top of adaptive learning, an agent may also have a set of possible models because of insufficient histories of data. A robust model needs to take this model uncertainty into account. In this paper, we consider model uncertainty regarding *physical dynamics* from two perspectives, both of which are closely related the predictability.[15] The first issue is the speed of learning. We can specify different values for the gain parameter γ, which controls the time variability of regression coefficients. A model with a small gain parameter would not be sensitive to new information, which means the agent slowly learns about structural changes. In an extreme case when γ is set to zero, the model boils down to a constant-coefficient case so the agent assumes there would not be structural breaks. The second issue we are concerned with regards the restrictions on the physical dynamics, which corresponds to the persistence of pricing factors. As Duffee (2011*a*) and Joslin, Priebsch and Singleton (2014) suggest high persistence may boost the predictive performance, we incorporate this point in a time-varying manner.

In a time-varying framework, when implementing joint estimation of coefficients and model probabilities for $k = 1, ..., K$ models, it means that we need to estimate the following sum:

$$p(\beta_{t-1}|D_{t-1}) = \sum_{k=1}^{K} p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1}) \mathbf{Pr}(L_{t-1} = k|D_{t-1}), \tag{4.19}$$

where $L_{t-1} = k$ means the k_{th} model is selected at time $t-1$ and $p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1})$ is given by the Kalman filter. Technical details regarding the computation of the above quantities are left to Appendix C.2.4 and are explained in detail in Koop and Korobilis (2012, 2013). We implement a dynamic model selection (DMS) approach that chooses the best model with the highest probability at any point in time, in order to obtain the optimal restrictions the representative agent imposes in a

[14] We also need the gain to be sufficiently small in order to ensure convergence, see Evans and Honkapohja (2001) or more technically, Benveniste, Métivier and Priouret (1990).

[15] In light of the argument in Joslin, Singleton and Zhu (2011), we focus on physical dynamics only as the predictive performance of pricing factors is unrelated to pricing dynamics in our setup.

time-varying manner.

4.2.2.3 Decomposition of the Sources of Uncertainty

Following Dangl and Halling (2012), we conduct the following variance decomposition from the law of total variance. Through the decomposition, we aim to understand all possible sources of uncertainty with respect to the prediction of our pricing factor $\mathcal{P}$.

Firstly, we can decompose the variance with respect to different choices of learning gain parameter γ:

$$Var(\mathcal{P}) = E_{\gamma}\big(Var(\mathcal{P}|\gamma)\big) + Var_{\gamma}\big(E(\mathcal{P}|\gamma)\big), \tag{4.20}$$

where the operators $E_{\gamma}(\cdot)$ and $Var_{\gamma}(\cdot)$ are the expectation and variance with regards to γ, respectively. The former term in the above equation can be further decomposed with respect to different choices of forecasting model L:

$$E_{\gamma}\big(Var(\mathcal{P}|\gamma)\big) = E_L\big(Var(\mathcal{P}|L,\gamma)\big) + Var_L\big(E(\mathcal{P}|L,\gamma)\big). \tag{4.21}$$

After some algebra and using the expressions detailed in previous sections and Appendix C.2, we have

$$\begin{aligned} Var(\mathcal{P}_{t+1}) \quad = \quad & \underbrace{\sum_j \big[\sum_k (\Sigma_t|L_k,\gamma_j,D_t)P(L_k|\gamma_j,D_t)\big]P(\gamma_j|D_t)}_{\text{Observational variance}} \\ & + \underbrace{\sum_j \big[\sum_k (X_t\Phi_{t|t-1}X_t'|L_k,\gamma_j,D_t)P(L_k|\gamma_j,D_t)\big]P(\gamma_j|D_t)}_{\text{Parameter uncertainty}} \\ & + \underbrace{\sum_j \big[\sum_k (\hat{\mathcal{P}}^j_{t+1,k} - \hat{\mathcal{P}}^j_{t+1})^2 P(L_k|\gamma_j,D_t)\big]P(\gamma_j|D_t)}_{\text{Model restriction uncertainty}} \\ & + \underbrace{\sum_j (\hat{\mathcal{P}}^j_{t+1} - \hat{\mathcal{P}}_{t+1})^2 P(\gamma_j|D_t)}_{\text{Learning speed uncertainty}}, \end{aligned} \tag{4.22}$$

where Σ_t denotes the variance of the disturbance term in the observation equation, $\Phi_{t|t-1}$ denotes the unconditional variance of the time-t prior of the coefficient vector β_t, $\hat{\mathcal{P}}^j_{t+1}$ is the weighted average conditional on γ_j and $\hat{\mathcal{P}}_{t+1}$ is the weighted average over all candidate models.

The individual terms of Equation (4.22) state the sources of prediction uncertainty and have intuitive interpretations. The first term measures the expected observational variance, calculated over different choices of learning gain γ and forecast model L. This term in fact captures the random fluctuations or risks in the pricing factors, relative to the predictable drift component. The second term is the expected variance from errors in the estimation of the coefficient vector, which can be interpreted as the source of estimation or parameter uncertainty. The third term captures model uncertainty with respect to model restrictions. The last term measures the uncertainty with respect to the learning speed, which can also be considered as the time variability of the model coefficients.

4.2.3 Portfolio Allocation under Uncertainty

In the last section we describe the term structure pricing model allowing for parameter and model uncertainty, but the uncertainty is not priced for the representative agent. That is to say, no matter how many models are available, provided a model estimated and selected by the agent ex post, there is no uncertainty but only interest-rate or inflation risk.[16] Investors may rebalance the portfolio because of speculation or hedging demand, but it is hard to tell whether the term premia accounts for the uncertainty or not, and to what degree. In the case where the representative agent truly requires compensation for the uncertainty, the market prices of risk may be overestimated.[17] Therefore, the model that does not explicitly take uncertainty premia into account can cause some anomalies, for example, high Sharpe ratios suggested by Duffee (2010). This can be explained by the inability to separate the uncertainty premia from the risk premia, see Knight (1921). We do not intend to decompose the term premia into risk premia and uncertainty premia in this paper, but we are interested in whether allowing for uncertainty aversion can increase economic value for a small short-term investor with no market impact.

The aversion to uncertainty is essential when we consider a short-term investment by holding a long-term bond for a relatively short period say one year, as Sangvinatsos and Wachter (2005) and Johannes, Korteweg and Polson (2013) suggest that failing to hedge out the uncertainty carries a high utility cost. A classical, or maybe naive, short-term investor who is given only one pricing model and who does not consider parameter uncertainty, can end up with an investment strategy with high volatility and has little economic value. In contrast, a sophisticated mean-variance investor will consider a robust strategy because he or she is averse to parameter and model uncertainty.[18]

4.2.3.1 A Mean-Variance Portfolio Model with Parameter Uncertainty Aversion

To begin with, we consider the classical mean-variance model proposed by Markowitz (1952) and Sharpe (1970), where the optimal portfolio weight of M^r risky assets, w, is given by the solution of

[16] This implies long-term investors do not perceive high uncertainty, because once an investment decision is made, they do not rebalance the portfolio frequently. Sangvinatsos and Wachter (2005) show that investors with long investment horizons indeed take extreme long positions in long-term bonds because of hedging demands. If a bond is held to maturity, the expected return is fixed and irrelevant to the model, given that the U.S. treasury bonds are usually considered non-defaultable. The long-term institutional investors hold major share of the U.S. bond market and hence has high market power.

[17] This in fact is a small-sample problem, which can be resolved with very long histories of data, as we can recover the true model with learning and dynamic model selection.

[18] Gargano, Pettenuzzo and Timmermann (2014) analyze the portfolio selection problem under uncertainty with power utility, but they do not consider robust control. Johannes, Korteweg and Polson (2013) suggest mean-variance utility is similar to power utility in absence of fat tails, so in this paper we only consider investors with mean-variance utility for simplicity.

the following optimization problem:[19]

$$\max_{w} w^\intercal \mu - \frac{\gamma}{2} w^\intercal \Sigma w, \tag{4.23}$$

where μ is the M^r-vector of the true expected excess returns over the risk-free asset, Σ is the $M^r \times M^r$ covariance matrix of excess returns, and the scalar γ is the risk aversion parameter. The solution to this problem is

$$w = \frac{1}{\gamma} \Sigma^{-1} \mu. \tag{4.24}$$

However, an investor knows that the expected excess returns are from a model which may generate imprecise estimates of expected excess returns $\hat{\mu}$, and therefore, pursues robustness when determining the portfolio weights. The demand for robustness is equivalent to investors' aversion to the uncertainty associated with the parameters estimated, see Gilboa and Schmeidler (1989) and Chen and Epstein (2002). To explicitly account for the uncertainty aversion, we introduce two elements to the above optimization problem following Garlappi, Uppal and Wang (2007). Firstly, the investor recognizes that the expected excess return for each asset can lie within a specified interval of its estimated value. This implies that the point estimate of the expected excess return is not the only possible value considered by the investor. Secondly, we introduce an additional optimization: The investor minimizes over the choice of expected returns, subject to the constraint of the specified interval.

The max-min problem above originates from the model of Gilboa and Schmeidler (1989), which is given by the form

$$\max_{w} \min_{\mu} w^\intercal \mu - \frac{\gamma}{2} w^\intercal \Sigma w, \tag{4.25}$$

subject to

$$f_C(\mu, \hat{\mu}, \Sigma) \leq \varepsilon. \tag{4.26}$$

To clarify the constraint (4.26), consider a case where the excess returns follow a multivariate Gaussian distribution with the true mean μ and the expected returns $\hat{\mu}$ are estimated by the sample mean with T observations. Then the quantity

$$T(\hat{\mu} - \mu)^\intercal \Sigma^{-1} (\hat{\mu} - \mu)$$

has a χ^2 distribution with M^r degree of freedom, where M^r is the dimension of the vector of returns.[20] Let $f_C = T(\hat{\mu} - \mu)^\intercal \Sigma^{-1} (\hat{\mu} - \mu)$ and ε be a chosen quantile for the χ^2 distribution. The constraint (4.26) can be expressed as

$$T(\hat{\mu} - \mu)^\intercal \Sigma^{-1} (\hat{\mu} - \mu) \leq \varepsilon.$$

[19] In order to keep the classical representation, the following equations in this section have abuse of notation γ. Note that the notation $\boldsymbol{\gamma}$ in bold in this section means the risk aversion parameter, which is different from the learning gain parameter γ in previous sections.

[20] If Σ is not known, then the quantity $\frac{T(T-M^r)}{(T-1)M^r}(\hat{\mu} - \mu)^\intercal \hat{\Sigma}^{-1} (\hat{\mu} - \mu)$ follows an F distribution with M^r and $T - M^r$ degrees of freedom, see Garlappi, Uppal and Wang (2007).

It means the constraint corresponds to a confidence interval in the probabilistic statement

$$P[T(\hat{\mu}-\mu)^{\mathsf{T}}\Sigma^{-1}(\hat{\mu}-\mu)\leq\varepsilon]=1-p,$$

for an appropriate level p, say 5%.

Now we parameterize the constraint (4.26) as

$$(\hat{\mu}-\mu)^{\mathsf{T}}\Sigma^{-1}(\hat{\mu}-\mu)\leq\varepsilon. \tag{4.27}$$

Then the max-min problem (4.25) subject to constraint (4.27) can be simplified into a maximization problem and we can obtain an intuitive expression of the optimal portfolio weights. Garlappi, Uppal and Wang (2007) have the following proposition:

Proposition 1 *The max-min problem (4.25) subject to constraint (4.27) is equivalent to the following maximization problem*

$$\max_{w} w^{\mathsf{T}}\hat{\mu}-\frac{\gamma}{2}w^{\mathsf{T}}\Sigma w-\sqrt{\varepsilon w^{\mathsf{T}}\Sigma w}. \tag{4.28}$$

The optimal portfolio weights for this problem can be expressed as

$$w^{*}=\frac{1}{\gamma}\Sigma^{-1}\Big(\frac{1}{1+\frac{\sqrt{\varepsilon}}{\gamma\sigma_p^{*}}}\Big)\hat{\mu}, \tag{4.29}$$

where σ_p^* is the variance of the optimal portfolio that can be obtained from solving a second degree polynomial equation, see Appendix C.3.1.

This framework nests the classical mean-variance portfolio. When $\varepsilon\to 0$, we immediately obtain Equation (4.24). Without loss of generality, the framework of parameter uncertainty aversion measures the effect of uncertainty aversion with respect to rare events, as indicated in Liu, Pan and Wang (2005). This means investors have robust control for rare events and allow for the worst-cast scenario that rare disasters actually happen.[21] With higher value of ε investors become more pessimistic. When $\varepsilon\to\infty$, investors become so pessimistic that they would not invest on any risky assets, which in turn gives a minimum-variance portfolio.

4.2.3.2 An Extension with Model Uncertainty

In this section, we extend the optimization problem (4.25) to characterize model uncertainty. In Equation (4.25) we only use one model to generate expected excess returns $\hat{\mu}$, without the freedom of selecting alternative models. Suppose we have a set of candidate models, then the max-min problem becomes

$$\max_{w,\hat{\mu}}\min_{\mu} w^{\mathsf{T}}\mu-\frac{\gamma}{2}w^{\mathsf{T}}\Sigma w, \tag{4.30}$$

[21] To see this, recall that ε has a probabilistic interpretation. Our max-min problem mimics investors' 'pessimism' that they assume the worst-case scenario will always happen when making investment decisions.

subject to

$$(\hat{\mu}-\mu)^{\intercal}\Sigma^{-1}(\hat{\mu}-\mu) \leq \varepsilon, \tag{4.31}$$

$$\hat{\mu} \in \{\hat{\mu}_k : k = 1, ..., K\}, \tag{4.32}$$

where $\hat{\mu}$ can be chosen from a set of K candidate models. The above max-min problem can also be simplified into a maximization problem which is easier to solve. Extending the results of Garlappi, Uppal and Wang (2007), we have the following proposition:

Proposition 2 *The max-min problem (4.30) subject to constraint (4.31) and (4.32) is equivalent to the following maximization problem*

$$\max_{w,\hat{\mu}} w^{\intercal}\hat{\mu} - \frac{\gamma}{2}w^{\intercal}\Sigma w - \sqrt{\varepsilon w^{\intercal}\Sigma w}, \tag{4.33}$$

subject to $\hat{\mu} \in \{\hat{\mu}_k : k = 1, ..., K\}$. *The optimal portfolio weights for this problem can be expressed as*

$$w^{*}* = \frac{1}{\gamma}\Sigma^{-1}\Big(\frac{1}{1+\frac{\sqrt{\varepsilon}}{\gamma\sigma_p^{*}}}\Big)\hat{\mu}^{*}, \tag{4.34}$$

where $\hat{\mu}^{*}$ *is the optimal expected excess return selected from K candidate models, and* σ_p^{*} *is the variance of the optimal portfolio that can be obtained from solving a second degree polynomial equation, see Appendix C.3.2.*

We can see how this extension contributes to investors' portfolio allocation in an intuitive way. Equation (4.31) shows we expand the feasible region in our minimum optimization problem, which is the same as the case of parameter uncertainty aversion. However, with condition (4.32), we then shrink the admissible region to the area associated with the optimal forecasts generated from candidate models. Indeed the region we search is expanded for the minimum optimization, but we only select the weights that give us higher utility in the maximum optimization. This refinement helps investors avoid unrealistic pessimistic investment, especially during the period when the minimum-variance strategy performs poorly. Even in a conservative situation where ε is large, investors still intend to hold some risky assets. It is indeed more realistic: An investor can hedge out risks by diversification according to a selected forecasting model, instead of being extremely pessimistic and only invest in risk-free assets.

4.3 Results

In this paper, we use the smoothed US bond yields provided from the US Federal Reserve by Gürkaynak, Sack and Wright (2007). We also include the 3- and 6-month Treasury Bills (Secondary Market Rate) from St. Louis Federal Reserve Economic Data (FRED). The empirical analysis focuses on yields with 12 maturities of 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108 and 120 months. The full sample of end of the month yield data is from June 1961 to October 2014. Our training sample has 121 observations from the beginning, up to and including June 1971. We do not introduce any other variables in addition to bond yield data, because we aim to understand the predictability that purely comes from the information in the bond market. From a finance viewpoint, we aim to explore all sources of the prediction uncertainty and how investors can benefit from a robust model which takes these sources of uncertainty into account.[22]

4.3.1 Pricing Dynamics and Market Prices of Risk

In our pricing setup, we specify a parsimonious factor structure so that a few portfolio risk factors can effectively model the term structure. Three risk factors – *Level*, *Slope* and *Curvature* – can capture most of the variance of bond yields, see Nelson and Siegel (1987) and Litterman and Scheinkman (1991). In line with most of the literature, we also use these three pricing factors or portfolios to price bonds when specifying our model. The portfolio weights are fixed for consistency and tractability.[23] Following Joslin, Singleton and Zhu (2011), we assume our portfolio risk factors $\mathcal{P}$ are measured without errors. Given this assumption, the cross-sectional arbitrage-free restrictions are irrelevant to the conditional point forecasts of $\mathcal{P}$ under $\mathbb{P}$. Therefore, we can separately estimate pricing dynamics, provided that the covariance of the innovations of $\mathcal{P}$ is estimated from physical dynamics.

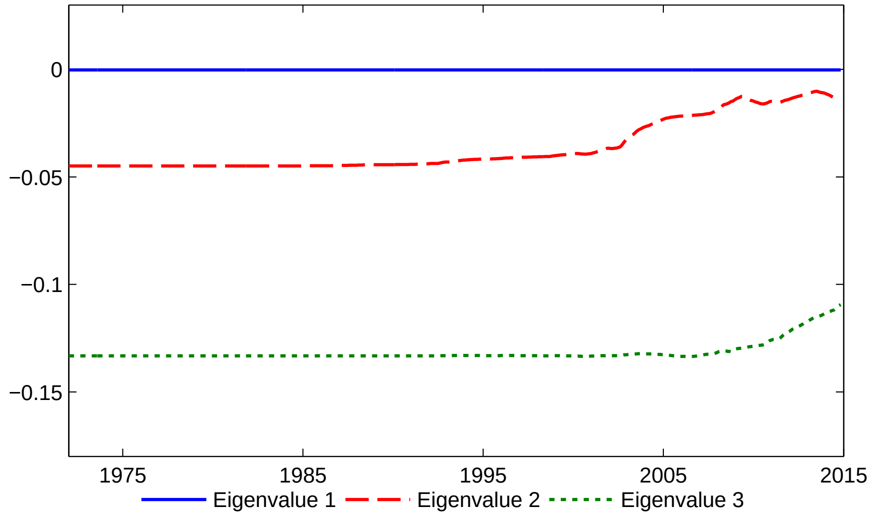
Figure 4.1 displays the evolution of three real eigenvalues $\lambda^{\mathbb{Q}}$ associated with $K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}$ over time. The first eigenvalue is slightly below zero, which implies that the Level factor is very persistent and close to a unit root process. The second eigenvalue is smaller but it still implies a highly persistent process of the Slope factor. The third eigenvalue is the smallest among the three, suggesting a less persistent evolution of the Curvature factor. We can see that the eigenvalues are stable over time, and therefore the arbitrage-free relation that specifies the link between market prices of risk and risk factors is unlikely to have a significant change. The above findings are consistent with Joslin, Priebsch and Singleton (2014) and Giacoletti, Laursen and Singleton (2014). Our new finding is that the factor process in pricing dynamics becomes more persistent in the recent decade, which implies a relatively

[22]Ludvigson and Ng (2009) and Joslin, Priebsch and Singleton (2014) suggest that unspanned macro information has predictive power for future movements of bond yields, whilst Bauer and Rudebusch (2015) provide evidence that some key macroeconomic variables are indeed spanned by bond yields. It is interesting to develop hybrid models with both spanned and unspanned macroeconomic risks, and explore the prediction uncertainty from different choices of predictors. We do not pursue this direction in this paper, although our framework can be easily extended to allow for the unspanned macro risks or hybrid models.

[23]Our findings are robust to different approximation methods of the portfolio weights and pricing factors.

more flat forward curve. The second eigenvalue is gradually growing up since the middle of the 2000s, while the third eigenvalue is rising from the start of the financial crisis.

Figure 4.1: *Eigenvalues* $\lambda^{\mathbb{Q}}$



Notes: This graph shows the estimates of time-varying parameters $\lambda^{\mathbb{Q}}$ associated with $K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}$, which govern the loading matrix $B^m_{\mathcal{P}}(\lambda^{\mathbb{Q}}_t)$ in Pricing Equation (4.13). Sample period is from 1971:07 to 2014:10.

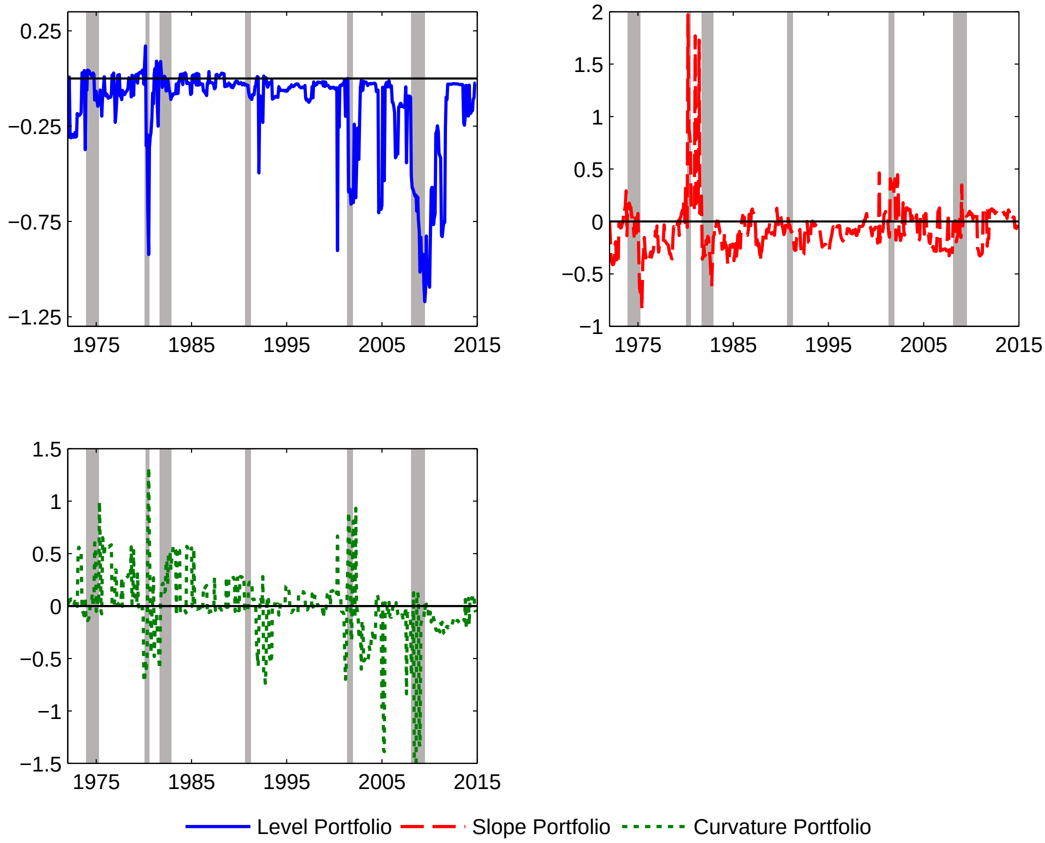
In this framework, we can gain knowledge about the priced risk by looking into Equation (4.7). Joslin, Priebsch and Singleton (2014) show that to a first-order approximation, the three elements of the scaled prices of risk represent the expected excess returns of three factor-mimicking portfolios, respectively. To be more specific, the excess return on a factor-mimicking portfolio, say Level, changes locally one-to-one with changes in the corresponding factor, but whose value is unresponsive to changes in other factors. Figure 4.2 depicts the one-period expected excess returns of Level, Slope and Curvature factor-mimicking portfolios.[24] Exposures to Level lose money if rates are expected to fall, which is usually when monetary policy is eased, for example, the recession periods. Exposures to Slope lose (gain) money if the curve is going to steepen (flatten), which is connected with the changes in the stance of monetary policy or investors' expectations, e.g. the monetarist experiment during the early 1980s. Exposures to Curvature may be difficult to interpreted, but Litterman, Scheinkman and Weiss (1991) link curvature to the volatility of the Level factor via the argument of yield curve convexity. We also find that the expected returns of Level portfolio heavily drop to historical low in the global financial crisis, which potentially reflects the 'flight-to-quality' demand suggested by Christensen, Lopez and Rudebusch (2010) and Bauer, Rudebusch and Wu (2014).[25]

Additionally, we can assess the economic significance of three pricing factors by calculating

[24]We relax the zero restrictions imposed on the price of the Curvature risk by Joslin, Priebsch and Singleton (2014) and Giacoletti, Laursen and Singleton (2014). This relaxation does not affect the power of in-sample fitting or out-of-sample forecasting in our framework. Further discussion is followed in the next section.

[25]Specifically, Bauer and Rudebusch (2015) indicate inflation measures are mainly correlated with the Level, 'measures of slack' are most closely correlated with the Slope, and growth measures are correlated most strongly with the Curvature. Similar evidence can also be found in Diebold and Rudebusch (2013).

Figure 4.2: *One-Period Expected Excess Returns of Factor-Mimicking Portfolios*



Notes:

1. This figure displays the one-period expected excess returns of Level, Slope and Curvature factor-mimicking portfolios from 1971:07 to 2014:10, which can be obtained from Equation (4.7).
2. Shaded areas are recession periods based on the NBER Recession Indicators. The unit is percentage.

their contribution to the variability of the pricing kernel. From Equation (4.2) we have

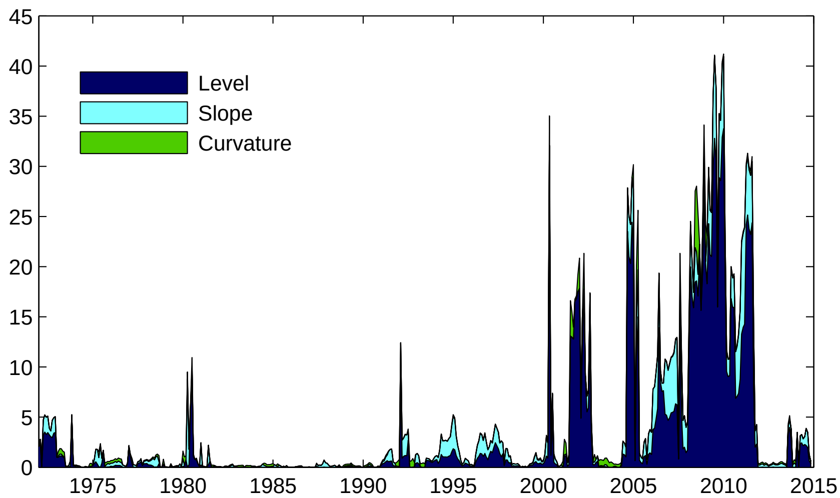
$$\ln \mathcal{M}_{\mathcal{P},t+1} = -r_t - \frac{1}{2}\Lambda'_{\mathcal{P}t}\Lambda_{\mathcal{P}t} - \Lambda'_{\mathcal{P}t}\varepsilon^{\mathbb{P}}_{\mathcal{P},t+1}. \tag{4.35}$$

Following Adrian, Crump and Moench (2013), we decompose the conditional volatility of the pricing kernel into the contributions due to each price of risk according to

$$Var(\ln \mathcal{M}_{\mathcal{P},t+1}) = \Lambda'_{\mathcal{P}t}\Lambda_{\mathcal{P}t} = \sum_{j=1}^{N} \Lambda^2_{j,\mathcal{P}t}. \tag{4.36}$$

Figure 4.3 sets out the contribution of risk prices of all three factors to the time variation of the pricing kernel. We find that there are several peaking periods of the time variance after 2000, and the time variance is extremely high around the financial crisis, which means agents' expectations of excess returns are very uncertain at that time. Consistent with Adrian, Crump and Moench (2013), pricing kernel time variance is largely, though not exclusively, driven by the Level risk, which implies that the expected excess returns on the long-term bonds are also largely driven by the Level risk.

Figure 4.3: *Pricing Kernel Variance Decomposition*



Notes: This graph shows the decomposition of the conditional volatility of the pricing kernel by decomposing $\Lambda'_{\mathcal{P}t}\Lambda_{\mathcal{P}t}$ into three components corresponding to Level, Slope and Curvature risks. Sample period is from 1971:07 to 2014:10.

4.3.2 Physical Dynamics and Out-of-Sample Predictability

As mentioned above, the physical dynamics are crucial for term structure predictability. In the model setup, the agent is able to learn about the evolution of parameters over time. Specifically, we allow for both time-varying volatility and coefficients, as Johannes, Korteweg and Polson (2013) and Gargano, Pettenuzzo and Timmermann (2014) suggest these extensions are useful in capturing time-varying features and improving out-of-sample predictability. In addition to parameter uncertainty, our method also consider model uncertainty and therefore is robust to potential structural breaks, see Gürkaynak and Wright (2012).

We introduce two layers of model uncertainty. The first layer is the learning speed, which is controlled by the learning gain parameter γ discussed in Section 4.2.2.1. We define a wide grid of values for γ: $[0, 0.01, 0.02, 0.03, 0.04]$, which covers the last two learning rules. While $\gamma = 0$ is equivalent to the constant parameter case, $\gamma = 0.04$ gives us a very flexible model as the observation two years ago only receive 45% as much weight as the newly incoming observation. The learning speed characterizes how agents adjust to structural changes and form their expectations, and hence is related to the out-of-sample predictability.

The next layer of model uncertainty is about the restrictions we impose on the physical dynamics. The restrictions are motivated by the finding of Joslin, Singleton and Zhu (2011) and Duffee (2011*a*) that cross-sectional restrictions are unrelated to the predictive performance.[26] Diebold and Li (2006) and Diebold, Rudebusch and Aruoba (2006) indicate that mixed evidence is found concerning the usefulness of various restrictions, where they consider both cases of related factors and unrelated factors. In our specification, we have in total 20 models, which nests the cases of related and unrelated factors in Diebold and Li (2006), as well as the random walk restrictions in Duffee (2011*a*).[27] Combining two layers of model uncertainty, we have in total $5 \cdot 20 = 100$ models for selection at each point in time.[28] Actually, our method is robust to model specification and can mitigate the small sample bias indicated in Duffee and Stanton (2012) and Bauer, Rudebusch and Wu (2012), as the one 'true' model will always be selected with a long history of data.

4.3.2.1 Learning about the (un)predictability in the term structure

As mentioned in previous sections, our proposed term structure model with learning nests most of the current term structure models using conditional information at each point in time. In terms of the predictive performance of bond yields, we can safely focus on the conditional forecasts of pricing factors only in our framework. Joslin, Singleton and Zhu (2011) have shown that the cross-sectional no-arbitrage restriction is irrelevant for the conditional forecast of $\mathcal{P}$ under measure $\mathbb{P}$.

In this section, we compare the out-of-sample performance of the proposed model with two

[26] Note that Joslin, Priebsch and Singleton (2014) impose restrictions on the market prices of risk which increase the persistence of the physical dynamics, and hence the out-of sample forecasting performance can be improved. Our flexible model selection framework in fact implicitly nests the restrictions of the same purpose, and therefore it is not necessary to explicitly impose zero restrictions on the market prices of risk.

[27] In addition to the unrestricted model, we restrict that the Level factor is unaffected by other two factors, so we have two zero restrictions in the first row; we further have $2^4 = 16$ combinations of zero restrictions imposed on off-diagonal elements of the remaining two rows of the coefficient matrix. We then have additionally three more models that ensure the first one, two and three factors follow random walks, respectively. Intuitively, these restrictions can enforce a high degree of persistence under $\mathbb{P}$ and hence may increase the forecast performance as suggested by Joslin, Priebsch and Singleton (2014).

[28] To speed up the estimation process, we employ the algorithm proposed by Koop and Korobilis (2013), see Appendix C.2 for technical details.

challenging benchmark models: random walk and the full-sample estimation following Joslin, Singleton and Zhu (2011).[29] Duffee (2002) remarks that it is hard for term structure models to beat the random walk, though the random walk cannot provide informative economic implications in terms of the dynamics of risk premia. The full-sample estimation of Joslin, Singleton and Zhu (2011) in fact is an in-sample forecasting exercise, which gains huge advantages as it incorporate the information of realized values. However, the full-sample estimation may be contaminated by the realized expectations of interest rates, which therefore do not perfectly reflect real-time expectations using conditional information.

Table 4.1 shows the predictive performance of the proposed model relative to benchmarks. The performance of the learning model is similar to the benchmark models, at least at shorter forecast horizons; the learning model even outperforms the benchmark models for some maturities. This is not surprising as conditional information should be helpful for short-horizon forecasts. Moreover, the short-rate forecasts, which is most related to future short rate expectations and term premium estimates, are relatively accurate even for longer horizons. Therefore, we can consider the term premium estimates of the learning model a plausible alternative to that of the model using full-sample information. However, a rather surprising result is that the performance of either the learning model or the benchmark using full-sample information, is close to that of random walk. This observation urges us to have a deeper understanding of the interest rate (un)predictability.

We explore the sources of prediction uncertainty by the variance decomposition noted in Section 4.2.2.3. The total prediction variance can be decomposed into observational variance, variance due to errors in the estimation of the coefficients (parameter uncertainty), variance due to model uncertainty in terms of the choice of the restrictions (restriction uncertainty), and variance due to model uncertainty regarding the choice of learning speed (learning speed uncertainty). Figure 4.4 displays the variance decomposition for three pricing factors, where Panel A shows that the predominant source of uncertainty is observational variance. This finding is consistent with the findings of Dangl and Halling (2012), as the asset prices frequently fluctuate randomly over their expected values. These fluctuations serve as the source of risk premia, and dominate the drift components in the term structure model. Therefore the fluctuations in fact contaminate the predictability of term structure models, especially during the periods when pricing factors are highly persistent.[30]

In Panel B of Figure 4.4, by excluding the observational variance we can focus upon the relative weights of the remaining sources of prediction uncertainty. The parameter uncertainty turns out to be the dominant source of prediction uncertainty, which implies parameter instability is another main source causing interest rate unpredictability. Therefore, a success forecasting model should at least consider the feature of time-varying parameters. The restriction uncertainty is less important but can be meaningful during certain periods. For example, restriction uncertainty rises steeply around

[29]The model of Joslin, Singleton and Zhu (2011) is in fact nested within our framework if we focus on out-of-sample performance.

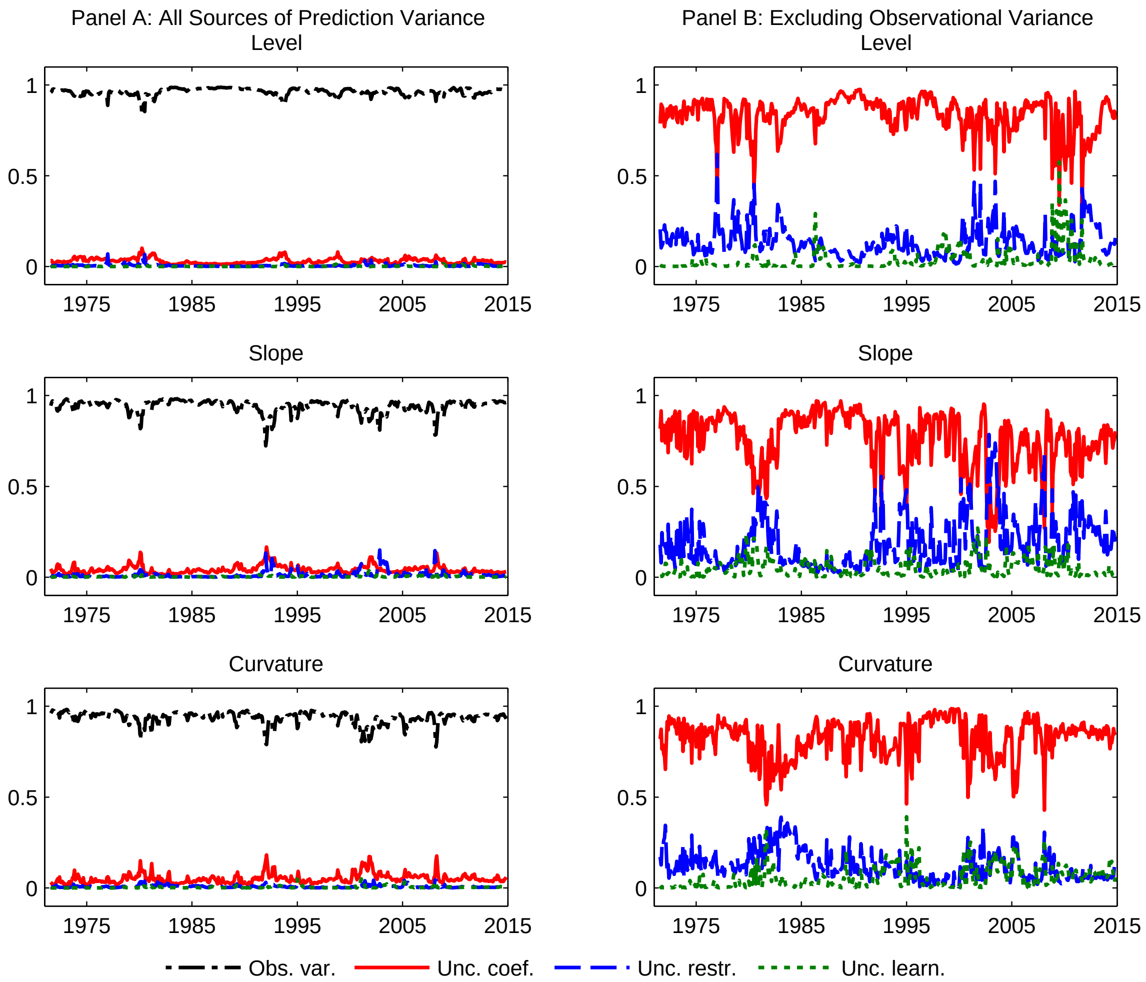
[30]This does not at all mean term structure models are not useful. For instance, term structure models can reveal informative dynamics of market prices of risks and have reliable term premia of long-term bonds, which can not be offered by the random walk model.

Table 4.1: *Predictive Performance of the Learning Model Relative to Benchmarks*

		FH \MA	3	6	12	24	36	48	60	72	84	96	108	120
MSPE	Learning versus RW	h=1	1.00	1.02	1.01	0.99	1.00	1.02	1.03	1.02	1.00	0.98	0.99	1.03
		h=3	0.99	1.01	0.98	1.00	1.01	1.02	1.02	1.02	1.01	1.01	1.01	1.02
		h=6	0.98	1.01	0.98	1.01	1.02	1.03	1.03	1.03	1.02	1.02	1.02	1.02
	Learning versus Full Sample	h=1	0.99	1.00	0.98	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.02
		h=3	1.00	1.01	0.99	1.01	1.01	1.02	1.02	1.03	1.03	1.03	1.04	1.05
		h=6	1.02	1.02	0.99	1.02	1.03	1.04	1.05	1.05	1.06	1.06	1.06	1.07
MAPE	Learning versus RW	h=1	1.01	1.03	1.06	1.01	1.01	1.02	1.02	1.02	1.01	1.01	1.01	1.02
		h=3	0.99	1.02	1.02	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02
		h=6	1.00	1.02	1.01	1.01	1.00	1.00	1.01	1.01	1.01	1.01	1.01	1.01
	Learning versus Full Sample	h=1	1.00	1.01	1.00	1.00	1.00	1.01	1.01	1.02	1.01	1.01	1.01	1.02
		h=3	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.03
		h=6	1.02	1.02	1.01	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.03	1.03

Notes:
1. This table shows 1-, 3- and 6-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from 1971:07 to 2014:10.
2. The MSPE-based and MAPE-based statistics relative to the random walk and full-sample estimation are reported.
3. In this table, we use following abbreviations. MSPE: mean squared prediction error; MAPE: mean absolute prediction error; RW: random walk; Full Sample: the full-sample (1961:06-2014:10) estimation following Joslin, Singleton and Zhu (2011); MA: maturity; FH: forecast horizon.

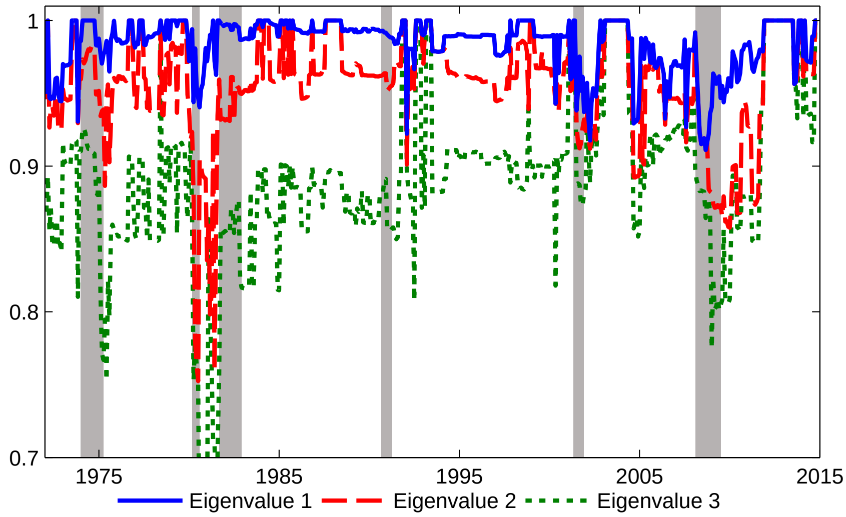
Figure 4.4: *Sources of Prediction Variance*



Notes:
1. This figure displays the decomposition of the prediction variance with respect to different sources.
2. In Panel A, the prediction variance is split into observational variance (Obs. var.), variance caused by errors in the estimation of coefficients (Unv. coef.), variance caused by the uncertainty with respect to the choice of restrictions (Unc. restr.) and variance caused by the uncertainty with respect to the learning speed (Unc. learn.). The illustration shows the relative weights of these components.
3. Panel B masks out observational variance and shows relative weights of the remaining variance.

the year 2003 and in the beginning of the financial crisis for the Slope factor. The uncertainty in learning speed is detectable but not of essence for most of the time. To highlight the importance of parameter uncertainty, Figure 4.5 sets out the persistence of the physical factor dynamics over time, which is examined by considering the behavior of the eigenvalues. There is a rising trend for the third eigenvalue since 1980s. We also detect significant drops in the persistence during recession periods, when the restrictions aiming to increase the persistence may not be valid.

Figure 4.5: *Eigenvalues* $\lambda^{\mathbb{P}}$



Notes: This graph shows the time-varying eigenvalues of estimated $K^{\mathbb{P}}_{t,\mathcal{PP}}$ in Eq (4.10). Sample period is from 1971:07 to 2014:10. Shaded areas are recession periods based on the NBER Recession Indicators.

Therefore, the large observational variance together with the high persistence in the data-generating process of bond pricing factors, gives rise to the similarity in the predictive performance between valid term structure models and the random walk. The inability to beat the random walk does not mean no predictability in excess returns, as excess return predictability is about whether excess returns can be explained by any pricing factors. The random walk model can be viewed as a special case of term structure models in which pricing factors are extremely persistent, and in that case the excess returns can be predicted by the pricing factors.

Campbell and Shiller (1991) indicate that no predictability in excess bond returns is equivalent to the Expectations Hypothesis (EH), and Adrian, Crump and Moench (2013) show the realized excess returns can be decomposed into innovations and a predictable element. If innovations are at a reasonable level, we should be able to detect predictability in excess returns by capturing the factor dynamics. Actually, Cieslak and Povala (2015*a*) show term premia is spanned by pricing factors and excess returns are predictable when compared with the EH benchmark.[31] However, in the previous literature, it seems difficult to translate the predictability of term premia into significant economic value, see for example, Della Corte, Sarno and Thornton (2008). In the following sections, we will

[31]The violation of the Expectations Hypothesis (EH) does not depend on the persistence of pricing factors, and hence the random walk model can also generate predictable excess returns if the loadings for short rates are not consistent with that for long rates.

evaluate that whether allowing for different sources of uncertainty can contribute to economic gains over the EH when investors make portfolio allocations.

4.3.3 Portfolio Selection

4.3.3.1 Predictability of excess returns

A simple approach to modeling the term structure is the Expectations Hypothesis (EH) that expected future short rates explain long rates. Campbell and Shiller (1991) indicates the empirical evidence fails to justify the strong form of Expectations Hypothesis and the idea that long-term interest rate are simply determined by the average of current and future expected short-term rates. However, EH could be resuscitated in weak form allowing for a constant term premia, consistent with an upward sloping yield curve. Based on the weak form of the Expectations Hypothesis, the long-term yield is average of expected future short term rates $y_t(\tau)^{EH}$ plus a constant risk premium C^{EH}:

$$y_t(\tau) = y_t(\tau)^{EH} + C^{EH}, \tag{4.37}$$

where the Expectations Hypothesis (EH) consistent bond yield $y_t(\tau)^{EH}$ is given by:

$$y_t(\tau)^{EH} = \frac{1}{\tau}\sum_{i=0}^{\tau-1} E_t y_{t+i}(1), \tag{4.38}$$

where $y_t(\tau)$ is the yield at time t for a bond of τ-period maturity. That is to say, the EH consistent long yield is equal to the average of expected short yields $E_t y_{t+i}(1)$.

The belief in Expectations Hypothesis is closely related to investors' behavior. If the weak form of the Expectations Hypothesis holds, then risk premia is constant. In other words, we should not be able to predict the short-term returns in the future. If an investor believes that the Expectations Hypothesis does not hold and the term premium should be time-varying, then the investor can rely on a specific prediction model to guide his/her portfolio choice.

To show the above argument, we follow Cieslak and Povala (2015*a*) to decompose the *excess holding period return*. First, we define the *holding period return* as the return on buying a τ-year zero coupon bond at time t and then selling it, as a $(\tau - m)$-year zero coupon bond, at time $t+m$. This holding period return is given by:

$$HPR_{t+m}(\tau, m) = \frac{1}{m}[p_{t+m}(\tau - m) - p_t(\tau)] \tag{4.39}$$

where $p_t(\tau)$ is the log price of τ-year zero coupon bond at time t and $p_{t+m}(\tau - m)$ is the log price of $(\tau - m)$-year zero coupon bond at time $t+m$. The difference between holding period return and the

m-year continuously compounded short yield is the *excess holding period return*:

$$EXR_{t+m}(\tau,m) = HPR_{t+m}(\tau,m) - y_t(m). \tag{4.40}$$

Note that a general form of term premium is given by

$$TP_t(\tau) = y_t(\tau) - y_t(\tau)^{EH}. \tag{4.41}$$

where $TP_t(\tau) \equiv C^{EH}$ if the EH holds. We can rewrite Equation (4.41) by relating the term premium to the *excess holding period return*:

$$TP_t(\tau) = \frac{1}{\tau} E_t[\sum_{i=0}^{\tau-2} EXR_{t+1+i}(\tau,1)]. \tag{4.42}$$

where $E_t[\cdot]$ is the expectation operator. By the linearity of expectation, we can write the 1-period ahead expected *excess holding period return* as

$$E_t[EXR_{t+1}(\tau,1)] = -(\tau-1)E_t[TP_{t+1}(\tau-1)] + \tau\, TP_t(\tau). \tag{4.43}$$

Therefore, it is not difficult to see that under the weak form of the Expectations Hypothesis, the m-period ahead expected *excess holding period return* is a constant:

$$E_t[EXR^{EH}_{t+m}(\tau,m)] = C^{EH}_m, \tag{4.44}$$

where C^{EH}_m is associated with the constant risk premium C^{EH} and often approximated by the historical average, see Rapach, Strauss and Zhou (2009) and Thornton and Valente (2012). But risk, and hence the term premia, is unlikely to be constant while underlying variables are changing. Cochrane and Piazzesi (2008) construct a test by regressing the excess bond returns on the forward rates and show that the forward rates have significant predictive power. Similar evidence can be found in Duffee (2002), Cochrane and Piazzesi (2005), Sarno, Thornton and Valente (2007), Tang and Xia (2007) and Gürkaynak and Wright (2012). In the case where the term premia is time-varying, the m-period *excess holding period return* is denoted by $xp_{t,m}$:

$$xp_{t,m} = E_t[EXR_{t+m}(\tau,m)] = -(\tau-m)E_t[TP_{t+m}(\tau-m)] + \tau\, TP_t(\tau). \tag{4.45}$$

It is straightforward to obtain $xp_{t,m}$ using a prediction model.

Although the EH is rejected by strong statistical evidence, it is puzzling that such predictability could not bring an improvement in economic utility of mean-variance investors, see Della Corte, Sarno and Thornton (2008), Thornton and Valente (2012) and Sarno, Schneider and Wagner (2014). Gargano, Pettenuzzo and Timmermann (2014) show that for investors with power utility and accounting for estimation error and model uncertainty, the predictability can be translated into higher economic value. Building upon previous literature, we consider a general framework that considers ambiguity aversion and nests ordinary risk aversion. It allows us to see if investors have significant improvement in their realized utility when considering potential sources of uncertainty.

4.3.3.2 Economic value

The evaluation of out-of-sample predictability does not consider the risk borne by an investor. It raises the issue of economic value of a forecasting model, as statistical significance does not measure its economic significance. Here we evaluate whether the model predictability is sufficiently large to be of economic value to risk-averse, or more generally, ambiguity-averse investors. Following Campbell and Thompson (2008), Welch and Goyal (2008), and Rapach, Strauss and Zhou (2009), we assume each investor, who is small and hence with no market impact, chooses portfolio weights based on the return forecasts. In this paper, we allow the investor to be able to invest in 10 fixed-income assets: 1- to 10-year US bonds. We then calculate realized utility gains for a mean-variance investor on a real-time basis.

To demonstrate the evaluation of the above strategies, we firstly discuss the case of an Expectations Hypothesis (EH) investor. We can compute the average utility for the mean-variance investor with relative risk aversion parameter γ_R who allocates his or her portfolio monthly among all assets using forecasts of the excess returns based on the historical average. This exercise requires the investor to forecast the variance of excess returns. Following Campbell and Thompson (2008), we assume that the investor estimates the variance matrix $\hat{\Sigma}_{t+1}^{-1}$ using a 5-year rolling window using monthly data of excess annually returns. A mean-variance investor who forecasts the excess bond returns using the historical average $\bar{r}_{t+1}$ will decide at the end of period t to allocate the following share of his or her portfolio to securities in period $t+1$:

$$w_{0,t} = \frac{1}{\gamma_R}\hat{\Sigma}_{t+1}^{-1}\bar{r}_{t+1} \tag{4.46}$$

where $\hat{\sigma}_{t+1}^2$ is the 5-year rolling-window estimate of the variance of excess returns.[32]

Over the out-of-sample period, the average of the realized utility of the investor is given by

$$\hat{v}_0 = \hat{\mu}_0 - (\frac{1}{2})\gamma_R\hat{\sigma}_0^2 \tag{4.47}$$

where $\hat{\mu}_0$ and $\hat{\sigma}_0^2$ are respectively the sample mean and variance of the excess holding period returns on the benchmark portfolio of the EH investor, which is constructed using forecasts of the excess returns based on the historical average.

Similarly, we can calculate the average utility for the same investor, when his or her decision is made by using a model j to forecast the excess bond returns. As it is noted in Section 4.2.3, we can construct the share $w_{j,t}$ based on the forecasts of model j.

The resulting realized average utility level is

$$\hat{v}_j = \hat{\mu}_j - (\frac{1}{2})\gamma_R\hat{\sigma}_j^2 \tag{4.48}$$

[32]Following Campbell and Thompson (2008), Rapach, Strauss and Zhou (2009) and Thornton and Valente (2012), we constrain the portfolio weight on bonds to lie between -100% and 200% each month, so in Eq. (4.46) $w_{0,t} = -1$ ($w_{0,t} = 2$) if $w_{0,t} < -1$ ($w_{0,t} > 2$).

where $\hat{\mu}_j$ and $\hat{\sigma}_j^2$ are the sample mean and variance of the excess holding period returns on the portfolio indexed by j. The investor forms the portfolio j using forecasts of the excess returns of bonds according to the jth forecasting model.

We can compute the utility gain, or certainty equivalent return, as the difference between $\hat{v}_j$ in Eq. (4.48) and $\hat{v}_0$ Eq. (4.47)

$$\Delta = \hat{v}_j - \hat{v}_0. \tag{4.49}$$

The utility gain that is expressed in average annualized percentage return can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in a predictive model relative to the information in the historical term premia alone. We report results for risk aversion parameters $\gamma_R = 1, 3, 6$; the results are qualitatively similar for other reasonably values (ranging from 1 to 10).

4.3.3.3 Performance of portfolio choice with uncertainty aversion

In this paper, we evaluate economic gains of 5 strategies holding for one year, over the mean-variance portfolio based on the Expectations Hypothesis (EH). The strategies reported in Table 4.2 include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA & MU). In Table 4.2 we also report different degrees of parameter uncertainty level, representing different views of rare events.[33]

The results are rather surprising. Our proposed model has very significant economic value in contrast to the EH benchmark. The utility gain based on the proposed model is ranging from 4% to a remarkably high value 27%. The economic value of the learning portfolios with uncertainty aversion and minimum-variance portfolio peak when $\gamma_R = 6$. The minimum-variance strategy naturally performs well at high risk-averse level, so learning portfolios with uncertainty aversion also have favorable performance as they are shrunk toward the minimum-variance strategy. Panel B of Table 4.2 shows when short sales are not allowed, the EH strategy seems to perform much worse, so the economic gains of strategies allowing for uncertainty aversion become extremely high.

It is worth noting that the strategies we proposed have very consistent performance. At low risk-averse level, whilst the minimum-variance portfolio have a relatively small economic gain (0.46%), the proposed strategies still have $4-6\%$ utility gains. This is because we search the portfolio weights in the admissible region based on reliable forecasts, so we do not fall into the 'pessimism trap' that no investment in risky assets is made.

Figure 4.6 gives the cumulative sum of log returns generated by the above strategies over time,

[33] As we have discussed in Section 4.2.3.1, the quantity ε has a probability interpretation corresponding to an F distribution with degrees of freedoms N and $T-N$.

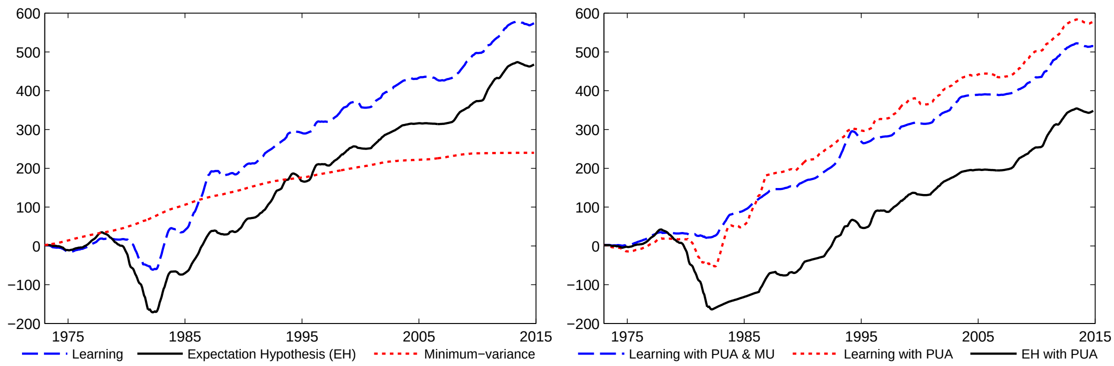
Table 4.2: *Economic Gains of Different Strategies*

Strategy		Utility gain (Δ)		
		$\gamma_R = 6$	$\gamma_R = 3$	$\gamma_R = 1$
Panel A: Short sales allowed				
	Minimum-variance	29.69	12.15	0.46
	Learning	11.76	7.14	4.08
	EH with PUA	3.86	0.52	-1.72
$\varepsilon = 2.78$ (99%)	Learning with PUA	13.93	8.28	4.50
	Learning with PUA&MU	27.22	13.12	6.18
	EH with PUA	0.63	0.07	-0.28
$\varepsilon = 2.07$ (95%)	Learning with PUA	12.77	7.51	4.01
	Learning with PUA&MU	27.05	13.26	6.23
Panel B: Short sales not allowed				
	Minimum-variance	94.43	39.87	3.49
	Learning	15.08	8.39	3.92
	EH with PUA	18.58	6.59	-1.40
$\varepsilon = 2.78$ (99%)	Learning with PUA	22.04	11.94	5.21
	Learning with PUA&MU	42.38	17.59	5.88
	EH with PUA	5.63	2.29	0.13
$\varepsilon = 2.07$ (95%)	Learning with PUA	18.40	9.77	4.02
	Learning with PUA&MU	42.12	17.29	5.67

Notes: 1. The table reports the out-of-sample 12-month holding utility gain (Δ) on different portfolio strategies, over the evaluation period from 1971:07 to 2014:10.
2. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences would be willing to pay, in order to switch from the Expectations Hypothesis (EH) strategy to another strategy. The utility gain is computed at three risk aversion levels, i.e., $\gamma_R = 1, 3, 6$. Higher utility gain is preferred.
3. We report the performance of strategies relative to the mean-variance portfolio based on the EH. The strategies reported include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU). In parenthesis, we report the percentage size of the confidence interval for an $F_{9,51}$ implied by the values of ε.

so we can have an intuitive impression. The minimum-variance portfolio has the smallest cumulative sum of log returns, but it is the most stable strategy which therefore can provide high economic value for a risk-averse investor. The learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU) is slightly less than the learning portfolio with parameter uncertainty only, but is more stable so it has a higher economic gain. The EH-based strategies perform less favorably owing to the drops in the late 1970s and early 1980s, probably because the economy was undergoing a structural change at that time.

Figure 4.6: *Cumulative Sum of Log Returns*



Notes:
1. This figure displays the cumulative sum of log returns generated by respective strategies when short sales are allowed and when we set $\gamma_R = 3$, $\varepsilon = 2.78$. The strategies include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU).
2. The evaluation period is from 1971:07 to 2014:10. The unit is percentage.

4.3.3.4 Robustness

In Figure 4.6 of the last section, we detect a notable fall in returns for the EH portfolio in and around 1980, while Federal Reserve's famous 'monetarist experiment' was conducted. We consider robustness checks by excluding this period. Table 4.3, 4.4 and 4.5 display the performance of the same strategies from 1990, 2000 and 2010 onward, respectively.

The resulting economic gains for our proposed portfolios are weakened, but still significant. Note that whilst the minimum-variance portfolios tend to have distinct performance at different risk-aver levels, the learning portfolios with uncertainty aversion perform stably and have significant positive gains (2% or more). Even when the minimum-variance portfolios have significantly negative gains, the performance of learning portfolios does not fall along the same way. Our proposed portfolios seem to have relatively weaker performance from 2000 onward, which we find is mainly due to highly uncertain estimates of pricing kernel in and around the financial crisis, recall Figure 4.3. This characteristic is potentially related to the decrease in persistence under the physical measure, see Figure 4.5. Moreover, the *zero lower bound* problem during the financial crisis may induce some unfeasible return forecasts which should be excluded when we construct the optimal portfolio.

After 2010, we see that the proposed portfolios get back on track and have economic value around 2%. Note that from these robustness checks, forecasts implied by our learning model alone does not guarantee substantial and consistent economic value, so ambiguity aversion is imperative in generating satisfactory economic utility. Therefore, by considering parameter and model uncertainty, investors truly benefit from the predictability of excess returns, and hence the *economic value puzzle* in bond returns can be resolved.

Moreover, by a simple utility gain decomposition from our results, we can reveal different degrees of utility gains because of aversion to various sources of uncertainty. The utility gain decomposition is done by computing the difference in gains among portfolios, i.e. learning, learning with PUA, and PUA&MU. The difference between learning portfolio and learning with PUA is the gain (if any) from the aversion to parameter uncertainty. The difference between learning portfolio and learning with PUA&MU is the gain from the ambiguity aversion allowing for both parameter and model uncertainty, which is generally much larger than the former one, except in very few cases. This finding is informative and confirms the necessity of incorporating model uncertainty. Although we have mentioned in Section 4.3.2.1 that the parameter uncertainty is the main source of prediction uncertainty when compared with model uncertainty, we find that *allowing for the smaller fraction of prediction variance originated from model uncertainty is of essence to generate significant and consistent economic value.* This finding further assures the robustness of our learning framework with model uncertainty.

Table 4.3: *Economic Gains of Different Strategies (from 1990 onward)*

	Strategy	Utility gain (Δ)		
		$\gamma_R = 6$	$\gamma_R = 3$	$\gamma_R = 1$
Panel A: Short sales allowed				
	Minimum-variance	1.21	-5.88	-10.60
	Learning	-0.93	-0.19	0.30
	EH with PUA	-0.56	-0.53	-0.51
$\varepsilon = 2.78$ (99%)	Learning with PUA	-0.89	-0.32	0.06
	Learning with PUA&MU	3.35	0.12	-0.44
	EH with PUA	-0.42	-0.37	-0.36
$\varepsilon = 2.07$ (95%)	Learning with PUA	-0.81	-0.21	0.18
	Learning with PUA&MU	3.22	0.25	-0.34
Panel B: Short sales not allowed				
	Minimum-variance	22.36	-2.86	-19.67
	Learning	-0.94	-0.29	0.15
	EH with PUA	-1.47	-1.27	-1.14
$\varepsilon = 2.78$ (99%)	Learning with PUA	-0.41	-0.29	-0.21
	Learning with PUA&MU	8.19	2.71	0.06
	EH with PUA	-1.09	-0.89	-0.76
$\varepsilon = 2.07$ (95%)	Learning with PUA	-0.35	-0.12	0.04
	Learning with PUA&MU	8.07	2.66	-0.06

Notes: 1. The table reports the out-of-sample 12-month holding utility gain (Δ) on different portfolio strategies, over the evaluation period from 1990:01 to 2014:10.
2. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences would be willing to pay, in order to switch from the Expectations Hypothesis (EH) strategy to another strategy. The utility gain is computed at three risk aversion levels, i.e., $\gamma_R = 1, 3, 6$. Higher utility gain is preferred.
3. We report the performance of strategies relative to the mean-variance portfolio based on the EH. The strategies reported include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU). In parenthesis, we report the percentage size of the confidence interval for an $F_{9,51}$ implied by the values of ε.

Table 4.4: *Economic Gains of Different Strategies (from 2000 onward)*

Strategy		Utility gain (Δ)		
		$\gamma_R = 6$	$\gamma_R = 3$	$\gamma_R = 1$
Panel A: Short sales allowed				
	Minimum-variance	-3.07	-8.03	-11.34
	Learning	-1.42	0.68	2.09
	EH with PUA	0.01	0.01	0.01
$\varepsilon = 2.78$ (99%)	Learning with PUA	-1.10	0.86	2.17
	Learning with PUA&MU	-0.50	-1.36	-1.41
	EH with PUA	0.01	0.01	0.00
$\varepsilon = 2.07$ (95%)	Learning with PUA	-0.99	0.95	2.25
	Learning with PUA&MU	-0.14	-1.40	-1.52
Panel B: Short sales not allowed				
	Minimum-variance	4.88	-12.70	-24.42
	Learning	-4.49	-1.31	0.80
	EH with PUA	0.00	0.00	0.00
$\varepsilon = 2.78$ (99%)	Learning with PUA	-3.84	-1.00	0.89
	Learning with PUA&MU	0.14	-0.43	0.28
	EH with PUA	0.00	0.00	0.00
$\varepsilon = 2.07$ (95%)	Learning with PUA	-3.47	-0.73	1.09
	Learning with PUA&MU	0.71	-0.38	0.28

Notes: 1. The table reports the out-of-sample 12-month holding utility gain (Δ) on different portfolio strategies, over the evaluation period from 2000:01 to 2014:10.
2. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences would be willing to pay, in order to switch from the Expectations Hypothesis (EH) strategy to another strategy. The utility gain is computed at three risk aversion levels, i.e., $\gamma_R = 1, 3, 6$. Higher utility gain is preferred.
3. We report the performance of strategies relative to the mean-variance portfolio based on the EH. The strategies reported include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU). In parenthesis, we report the percentage size of the confidence interval for an $F_{9,51}$ implied by the values of ε.

Table 4.5: *Economic Gains of Different Strategies (from 2010 onward)*

Strategy		Utility gain (Δ)		
		$\gamma_R = 6$	$\gamma_R = 3$	$\gamma_R = 1$
Panel A: Short sales allowed				
	Minimum-variance	-1.08	-7.74	-12.17
	Learning	1.06	0.61	0.30
	EH with PUA	0.00	0.00	0.00
$\varepsilon = 2.78$ (99%)	Learning with PUA	1.06	0.61	0.30
	Learning with PUA&MU	2.84	2.36	1.74
	EH with PUA	0.00	0.00	0.00
$\varepsilon = 2.07$ (95%)	Learning with PUA	1.06	0.61	0.30
	Learning with PUA&MU	2.35	2.10	1.74
Panel B: Short sales not allowed				
	Minimum-variance	14.85	-7.08	-21.70
	Learning	-0.06	0.02	0.07
	EH with PUA	0.00	0.00	0.00
$\varepsilon = 2.78$ (99%)	Learning with PUA	-0.06	0.02	0.07
	Learning with PUA&MU	1.82	0.75	0.33
	EH with PUA	0.00	0.00	0.00
$\varepsilon = 2.07$ (95%)	Learning with PUA	-0.06	0.02	0.07
	Learning with PUA&MU	1.82	0.75	0.79

Notes: 1. The table reports the out-of-sample 12-month holding utility gain (Δ) on different portfolio strategies, over the evaluation period from 2010:01 to 2014:10.
2. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences would be willing to pay, in order to switch from the Expectations Hypothesis (EH) strategy to another strategy. The utility gain is computed at three risk aversion levels, i.e., $\gamma_R = 1, 3, 6$. Higher utility gain is preferred.
3. We report the performance of strategies relative to the mean-variance portfolio based on the EH. The strategies reported include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU). In parenthesis, we report the percentage size of the confidence interval for an $F_{9,51}$ implied by the values of ε.

4.4 Conclusion

This paper studies the problem of a representative agent who learns about the information in the bond market over time, with the consideration of parameter uncertainty and model uncertainty. In addition to adaptive learning about parameters as considered in Giacoletti, Laursen and Singleton (2014), this proposed framework provides flexibility in specifying different learning speeds and model restrictions. The optimal specification can be selected according to predictive performance over time, and, therefore, reduce the risk of data snooping. This method is robust in the sense that it reveals the agent's expectations in real time by using conditional information. We find that apart from observational variance, parameter instability is the dominant driving force of predictive uncertainty, when compared with uncertainty in learning speed or model restrictions. It suggests that a successful term structure model should at least consider time-varying parameters when making conditional forecasts.

The problem of asset allocation for an investor with ambiguity aversion building upon Garlappi, Uppal and Wang (2007) is studied. After learning the parameters, the ambiguity-averse investor forms optimal portfolios by maximizing mean-variance expected utility. The ensemble of all salient features offered by our framework is essential in producing significant and consistent economic value over the Expectations Hypothesis benchmark. Ambiguity aversion with model uncertainty ensures that the search for portfolio weights is in a reliable region, which in turn not only boosts but also stabilizes the gains. Therefore, ambiguity aversion is a key to salvaging the models with significant predictability but little economic value used in the previous literature, and the *economic value puzzle* in bond returns can be resolved following this direction.

There are various important directions in which this approach can be extended. By allowing for more general model specifications, such as incorporating more information from macro-finance predictor variables or economic constraints as in Pettenuzzo, Timmermann and Valkanov (2014), it is possible to further improve model performance and provide meaningful economic rationales. It would also be interesting to develop hybrid models with both spanned and unspanned macroeconomic risks and explore the prediction uncertainty from different choices of predictors, as suggested by Bauer and Rudebusch (2015). Finally, our results suggest that the zero lower bound problem could hinder the performance of our portfolio strategy. We leave these directions for further research.

Bibliography

Abbritti, Mirko, Salvatore Dell'Erba, Antonio Moreno, and Sergio Sola. 2013. "Global factors in the term structure of interest rates." International Monetary Fund International Monetary Fund Working Paper WP/13/223.

Adrian, Tobias, Richard K. Crump, and Emanuel Moench. 2013. "Pricing the term structure with linear regressions." *Journal of Financial Economics*, 110(1): 110–138.

Altavilla, Carlo, Raffaella Giacomini, and Giuseppe Ragusa. 2014. "Anchoring the Yield Curve Using Survey Expectations." European Central Bank Working Paper Series 1632.

Andrews, Donald W.K., and Werner Ploberger. 1994. "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica*, 62(6): 1383–1414.

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

Anh, Le, and Scott Joslin. 2013. "Interest rate volatility and no-arbitrage affine term structure models." University of North Carolina at Chapel Hill Working Paper.

Avramov, Doron. 2002. "Stock return predictability and model uncertainty." *Journal of Financial Economics*, 64(3): 423–458.

Avramov, Doron, and Tarun Chordia. 2006. "Asset pricing models and financial market anomalies." *Review of Financial Studies*, 19(3): 1001–1040.

Bacchetta, Philippe, and Eric Van Wincoop. 2004. "A scapegoat model of exchange-rate fluctuations." *American Economic Review*, 94(2): 114–118.

Bai, Jushan, and Peng Wang. 2015. "Identification and Bayesian estimation of dynamic factor models." *Journal of Business and Economic Statistics*, 33(2): 221–240.

Bai, Jushan, and Serena Ng. 2006. "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions." *Econometrica*, 74(4): 1133–1150.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2013. "Measuring economic policy uncertainty." The University of Chicago Booth School of Business Working Paper.

Bansal, Naresh, Robert A. Connolly, and Chris Stivers. 2014. "The stock-bond return relation, the term-structure's slope, and asset-class risk dynamics." *Journal of Financial and Quantitative Analysis*, FirstView: 1–44.

Bansal, Ravi, and Ivan Shaliastovich. 2010. "Confidence risk and asset prices." *American Economic Review*, 100(2): 537–41.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler. 2005. "Comovement." *Journal of Financial Economics*, 75(2): 283–317.

Bauer, Gregory H., and Antonio Diez de los Rios. 2012. "An international dynamic term structure model with economic restrictions and unspanned risks." Bank of Canada Working Paper.

Bauer, Michael D. 2015. "Restrictions on risk prices in dynamic term structure models." Federal Reserve Bank of San Francisco Working Paper.

Bauer, Michael D., and Glenn D. Rudebusch. 2015. "Resolving the spanning puzzle in macro-finance term structure models." Federal Reserve Bank of San Francisco Working Paper.

Bauer, Michael D., and James D. Hamilton. 2015. "Robust bond risk premia." University of California at San Diego Working Paper.

Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu. 2012. "Correcting estimation bias in dynamic term structure models." *Journal of Business and Economic Statistics*, 30(3): 454–467.

Bauer, Michael D., Glenn D. Rudebusch, and Jing Cynthia Wu. 2014. "Comment on 'term premia and inflation uncertainty: empirical evidence from an international panel dataset'." *American Economic Review*, 104(1): 323–337.

Benhabib, Jess, and Pengfei Wang. 2015. "Private information and sunspots in sequential asset markets." *Journal of Economic Theory*, 158, Part B: 558 – 584.

Benveniste, Albert, Michel Métivier, and Pierre Priouret. 1990. *Adaptive Algorithms and Stochastic Approximations.* Springer.

Bernanke, Ben, Mark Gertler, and Simon Gilchrist. 1996. "The financial accelerator and the flight to quality." *The Review of Economics and Statistics*, 78(1): 1–15.

Bernanke, Ben S., and Jean Boivin. 2003. "Monetary policy in a data-rich environment." *Journal of Monetary Economics*, 50(3): 525–546.

Bianchi, Francesco, Haroon Mumtaz, and Paolo Surico. 2009. "The great moderation of the term structure of UK interest rates." *Journal of Monetary Economics*, 56(6): 856–871.

Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon. 2012. "Econometric measures of connectedness and systemic risk in the finance and insurance sectors." *Journal of Financial Economics*, 104(3): 535–559.

Björk, Tomas, and Bent Jesper Christensen. 1999. "Interest rate dynamics and consistent forward rate curves." *Mathematical Finance*, 9(4): 323–348.

Bloom, Nicholas. 2014. "Fluctuations in uncertainty." *The Journal of Economic Perspectives*, 28(2): 153–175.

Brandt, Michael W., Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud. 2005. "A simulation approach to dynamic portfolio choice with an application to learning about return predictability." *Review of Financial Studies*, 18(3): 831–873.

Byrne, Joseph P., Giorgio Fazio, and Norbert Fiess. 2012. "Interest rate co-movements, global factors and the long end of the term spread." *Journal of Banking and Finance*, 36(1): 183–192.

Cagetti, Marco, Lars Peter Hansen, Thomas Sargent, and Noah Williams. 2002. "Robustness and pricing with uncertain growth." *Review of Financial Studies*, 15(2): 363–404.

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

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

Carriero, Andrea, and Raffaella Giacomini. 2011. “How useful are no-arbitrage restrictions for forecasting the term structure of interest rates?” *Journal of Econometrics*, 164(1): 21–34.

Carriero, Andrea, George Kapetanios, and Massimiliano Marcellino. 2012. “Forecasting government bond yields with large Bayesian vector autoregressions.” *Journal of Banking & Finance*, 36(7): 2026–2047.

Carter, Chris K., and Robert Kohn. 1994. “On Gibbs sampling for state space models.” *Biometrika*, 81(3): 541–553.

Chen, Zengjing, and Larry Epstein. 2002. “Ambiguity, risk, and asset returns in continuous time.” *Econometrica*, 70(4): 1403–1443.

Christensen, Jens H.E., and Glenn D. Rudebusch. 2012. “The response of interest rates to US and UK quantitative easing.” *The Economic Journal*, 122(564): F385–F414.

Christensen, Jens H.E., Francis X. Diebold, and Glenn D. Rudebusch. 2011. “The affine arbitrage-free class of Nelson-Siegel term structure models.” *Journal of Econometrics*, 164(1): 4–20.

Christensen, Jens H.E., Jose A. Lopez, and Glenn D. Rudebusch. 2010. “Inflation expectations and risk premiums in an arbitrage-free model of nominal and real bond yields.” *Journal of Money, Credit and Banking*, 42(s1): 143–178.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. “Nominal rigidities and the dynamic effects of a shock to monetary policy.” *Journal of Political Economy*, 113(1): 1–45.

Cieslak, Anna, and Pavol Povala. 2014. “Expecting the fed.” Northwestern University Working Paper.

Cieslak, Anna, and Pavol Povala. 2015*a*. “Expected returns in treasury bonds.” *Review of Financial Studies*, 28(10): 2859–2901.

Cieslak, Anna, and Pavol Povala. 2015*b*. “Information in the term structure of yield curve volatility.” *Journal of Finance*, Forthcoming.

Cochrane, John H., and Monika Piazzesi. 2005. “Bond risk premia.” *American Economic Review*, 95(1): 138–160.

Cochrane, John H., and Monika Piazzesi. 2008. “Decomposing the yield curve.” Stanford University Working Paper.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A. Lochstoer. 2013. “Parameter learning in general equilibrium: The asset pricing implications.” National Bureau of Economic Research Working Paper.

Coroneo, Laura, Domenico Giannone, and Michele Modugno. 2015. “Unspanned macroeconomic factors in the yield curve.” *Journal of Business and Economic Statistics*.

Coroneo, Laura, Ken Nyholm, and Rositsa Vidova-Koleva. 2011. “How arbitrage-free is the Nelson-Siegel model?” *Journal of Empirical Finance*, 18(3): 393–407.

Cremers, K.J. Martijn. 2002. "Stock return predictability: A Bayesian model selection perspective." *Review of Financial Studies*, 15(4): 1223–1249.

D'Agostino, Antonello, Domenico Giannone, and Paolo Surico. 2006. "(Un)Predictability and macroeconomic stability." European Central Bank Working Paper Series 0605.

Dai, Qiang, and Kenneth J. Singleton. 2000. "Specification analysis of affine term structure models." *The Journal of Finance*, 55(5): 1943–1978.

Dai, Qiang, and Kenneth J. Singleton. 2003. "Term Structure Dynamics in Theory and Reality." *Review of Financial Studies*, 16(3): 631–678.

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

Della Corte, Pasquale, Lucio Sarno, and Daniel L. Thornton. 2008. "The expectation hypothesis of the term structure of very short-term rates: Statistical tests and economic value." *Journal of Financial Economics*, 89(1): 158–174.

Den Haan, Wouter J., and Steven W. Sumner. 2004. "The comovement between real activity and prices in the G7." *European Economic Review*, 48(6): 1333–1347.

Dewachter, Hans, and Leonardo Iania. 2012. "An extended macro-finance model with financial factors." *Journal of Financial and Quantitative Analysis*, 46(06): 1893–1916.

Dewachter, Hans, and Marco Lyrio. 2008. "Learning, macroeconomic dynamics and the term structure of interest rates." In *Asset Prices and Monetary Policy*. 191–245. University of Chicago Press.

Dewachter, Hans, Leonardo Iania, and Marco Lyrio. 2014. "Information in the yield curve: A Macro-Finance approach." *Journal of Applied Econometrics*, 29(1): 42–64.

Diebold, Francis X., and Canlin Li. 2006. "Forecasting the term structure of government bond yields." *Journal of Econometrics*, 130(2): 337–364.

Diebold, Francis X., and Glenn D. Rudebusch. 2013. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach. The Econometric and Tinbergen Institutes Lectures*, Princeton University Press.

Diebold, Francis X., and Kamil Yilmaz. 2009. "Measuring financial asset return and volatility spillovers, with application to global equity markets." *The Economic Journal*, 119(534): 158–171.

Diebold, Francis X, and Kamil Yilmaz. 2014. "On the network topology of variance decompositions: Measuring the connectedness of financial firms." *Journal of Econometrics*, 182(1): 119–134.

Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13(3): 253–263.

Diebold, Francis X., Canlin Li, and Vivian Z. Yue. 2008. "Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach." *Journal of Econometrics*, 146(2): 351–363.

Diebold, Francis X., Glenn D. Rudebusch, and S. Borağan Aruoba. 2006. "The macroeconomy and the yield curve: A dynamic latent factor approach." *Journal of Econometrics*, 131(1): 309–338.

Duffee, Gregory R. 2002. "Term premia and interest rate forecasts in affine models." *Journal of Finance*, 57(1): 405–443.

Duffee, Gregory R. 2010. “Sharpe ratios in term structure models.” Department of Economics, Johns Hopkins University Working Paper.

Duffee, Gregory R. 2011*a*. “Forecasting with the term structure: The role of no-arbitrage restrictions.” Johns Hopkins University, Department of Economics Working Paper.

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

Duffee, Gregory R. 2013. “Bond pricing and the macroeconomy.” In *Handbook of the Economics of Finance*. Vol. 2, Part B, , ed. George M. Constantinides, Milton Harris and Rene M. Stulz, 907–967. Elsevier.

Duffee, Gregory R. 2014. “Expected inflation and other determinants of Treasury yields.” Johns Hopkins University, Department of Economics Working Paper.

Duffee, Gregory R, and Richard H Stanton. 2012. “Estimation of dynamic term structure models.” *The Quarterly Journal of Finance*, 02(02): 1250008.

Duffie, Darrell. 2001. *Dynamic Asset Pricing Theory.* Princeton University Press.

Duffie, Darrell, and Rui Kan. 1996. “A yield-factor model of interest rates.” *Mathematical Finance*, 6(4): 379–406.

Eickmeier, Sandra, Leonardo Gambacorta, and Boris Hofmann. 2014. “Understanding global liquidity.” *European Economic Review*, 68: 1–18.

Elliott, Graham, and Allan Timmermann. 2008. “Economic forecasting.” *Journal of Economic Literature*, 46(1): 3–56.

Estrella, Arturo, and Frederic S. Mishkin. 1998. “Predicting US recessions: Financial variables as leading indicators.” *Review of Economics and Statistics*, 80(1): 45–61.

Evans, Charles L., and David A. Marshall. 2007. “Economic determinants of the nominal treasury yield curve.” *Journal of Monetary Economics*, 54(7): 1986–2003.

Evans, George W., and Seppo Honkapohja. 2001. *Learning and Expectations in Macroeconomics.* Princeton University Press.

Fama, Eugene F., and Robert R. Bliss. 1987. “The information in long-maturity forward rates.” *American Economic Review*, 77(4): 680–692.

Feng, Ziding D., and Charles E. McCulloch. 1996. “Using bootstrap likelihood ratios in finite mixture models.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(3): 609–617.

Fernández-Villaverde, Jesús, and Thomas J. Sargent Mark W. Watson Rubio-Ramírez, Juan F. 2007. “ABCs (and Ds) of Understanding VARs.” *The American Economic Review*, 97(3): 1021–1026.

Feunou, Bruno, Jean-Sébastien Fontaine, Anh Le, and Christian Lundblad. 2014. “Term structure modeling when monetary policy is unconventional: A new approach.” Bank of Canada Working Paper.

Filipović, Damir. 1999. “A note on the Nelson–Siegel family.” *Mathematical Finance*, 9(4): 349–359.

Gagliardini, Patrick, Paolo Porchia, and Fabio Trojani. 2009. "Ambiguity aversion and the term structure of interest rates." *Review of Financial Studies*, 22(10): 4157–4188.

Gargano, Antonio, Davide Pettenuzzo, and Allan G. Timmermann. 2014. "Bond return predictability: Economic value and links to the macroeconomy." University of California, San Diego Working Paper.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang. 2007. "Portfolio selection with parameter and model uncertainty: A multi-prior approach." *Review of Financial Studies*, 20(1): 41–81.

Geweke, John, and Gianni Amisano. 2010. "Comparing and evaluating Bayesian predictive distributions of asset returns." *International Journal of Forecasting*, 26(2): 216–230.

Giacoletti, Marco, Kristoffer T. Laursen, and Kenneth J. Singleton. 2014. "Learning, dispersion of beliefs, and risk premiums in an aribitrage-free term structure model." Stanford University Working Paper.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics*, 18(2): 141–153.

Gürkaynak, Refet S., and Jonathan H. Wright. 2012. "Macroeconomics and the term structure." *Journal of Economic Literature*, 50(2): 331–367.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright. 2007. "The US Treasury yield curve: 1961 to the present." *Journal of Monetary Economics*, 54(8): 2291–2304.

Hansen, Lars Peter. 2007. "Beliefs, doubts and learning: Valuing macroeconomic risk." *American Economic Review*, 97(2): 1–30.

Hanson, Bruce E. 2002. "Tests for parameter instability in regressions with I(1) processes." *Journal of Business and Economic Statistics*, 20(1): 45–59.

Hautsch, Nikolaus, and Fuyu Yang. 2012. "Bayesian inference in a stochastic volatility Nelson-Siegel model." *Computational Statistics and Data Analysis*, 56(11): 3774–3792.

Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho. 2011. "What factors drive global stock returns?" *Review of Financial Studies*, 24(8): 2527–2574.

Hubrich, Kirstin, Antonello D'Agostino, et al. 2013. "Financial shocks and the macroeconomy: Heterogeneity and non-linearities." European Central Bank ECB Occasional Paper Series 143.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. "Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software." *PlOS ONE*, 9(6): e98679.

Johannes, Michael, Arthur Korteweg, and Nicholas Polson. 2013. "Sequential learning, predictability, and optimal portfolio returns." *The Journal of Finance*, 69(2): 4.

Joslin, Scott, Anh Le, and Kenneth J. Singleton. 2013. "Gaussian macro-finance term structure models with lags." *Journal of Financial Econometrics*, 11(4): 581–609.

Joslin, Scott, Kenneth J. Singleton, and Haoxiang Zhu. 2011. "A new perspective on Gaussian dynamic term structure models." *Review of Financial Studies*, 24(3): 926–970.

Joslin, Scott, Marcel Priebsch, and Kenneth J. Singleton. 2014. "Risk premiums in dynamic term structure models with unspanned macro risks." *The Journal of Finance*, 69(3): 1197–1233.

Jotikasthira, Pab, Anh Le, and Christian T. Lundblad. 2015. "Why do term structures in different currencies comove?" *Journal of Financial Economics*, 115(1): 58–83.

Kim, Chang-Jin, and Charles R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* Vol. 1, the MIT Press.

Kim, Don H. 2009. "Challenges in Macro-Finance modeling." *Federal Reserve Bank of St. Louis Review*, 91(5, Part 2): 519–44.

Kim, Don H., and Jonathan H. Wright. 2005. "An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates." Board of Governors of the Federal Reserve System Working Paper.

Kim, Don H., and Kenneth J. Singleton. 2012. "Term structure models and the zero bound: An empirical investigation of Japanese yields." *Journal of Econometrics*, 170(1): 32–49.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit.* Hougton Mifflin Company.

Koop, Gary, and Dimitris Korobilis. 2009. "Bayesian multivariate time series methods for empirical macroeconomics." *Foundations and Trends® in Econometrics*, 3(4): 267–358.

Koop, Gary, and Dimitris Korobilis. 2012. "Forecasting inflation using Dynamic Model Averaging." *International Economic Review*, 53(3): 867–886.

Koop, Gary, and Dimitris Korobilis. 2013. "Large time-varying parameter VARs." *Journal of Econometrics*, 177(2): 185–198.

Koop, Gary, and Dimitris Korobilis. 2014. "A new index of financial conditions." *European Economic Review*, 71(0): 101 – 116.

Koop, Gary, M. Hashem Pesaran, and Simon M. Potter. 1996. "Impulse response analysis in nonlinear multivariate models." *Journal of Econometrics*, 74(1): 119–147.

Koopman, Siem Jan, Max I.P. Mallee, and Michel Van der Wel. 2010. "Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters." *Journal of Business and Economic Statistics*, 28(3): 329–343.

Kose, M. Ayhan, Christopher Otrok, and Charles H. Whiteman. 2003. "International business cycles: World, region, and country-specific factors." *American Economic Review*, 93(4): 1216–1239.

Kozicki, Sharon, and Peter A. Tinsley. 2001. "Shifting endpoints in the term structure of interest rates." *Journal of Monetary Economics*, 47(3): 613–652.

Krippner, Leo. 2015. "A theoretical foundation for the Nelson-Siegel class of yield curve models." *Journal of Applied Econometrics*, 30(1): 97–118.

Kumar, Alok, and Charles Lee. 2006. "Retail investor sentiment and return comovements." *The Journal of Finance*, 61(5): 2451–2486.

Kurmann, André, and Christopher Otrok. 2013. "News shocks and the slope of the term structure of interest rates." *American Economic Review*, 103(6): 2612–32.

Kurz, Mordecai. 1994. "On rational belief equilibria." *Economic Theory*, 4(6): 859–876.

Laubach, Thomas, Robert J. Tetlow, and John C. Williams. 2007. “Learning and the role of macroeconomic factors in the term structure of interest rates.” Board of Governors of the Federal Reserve System Working Paper.

Lee, Bong-Soo. 1998. “Permanent, temporary, and non-fundamental components of stock prices.” *Journal of Financial and Quantitative Analysis*, 33(01): 1–32.

Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang. 2013. “Fiscal foresight and information flows.” *Econometrica*, 81(3): 1115–1145.

Litterman, Robert B., and Jose Scheinkman. 1991. “Common factors affecting bond returns.” *The Journal of Fixed Income*, 1(1): 54–61.

Litterman, Robert B., José Scheinkman, and Laurence Weiss. 1991. “Volatility and the yield curve.” *The Journal of Fixed Income*, 1(1): 49–53.

Liu, Jun, Jun Pan, and Tan Wang. 2005. “An equilibrium model of rare-event premia and its implication for option smirks.” *Review of Financial Studies*, 18(1): 131–164.

Ludvigson, Sydney C. 2004. “Consumer confidence and consumer spending.” *The Journal of Economic Perspectives*, 18(2): 29–50.

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

Markowitz, Harry. 1952. “Portfolio Selection.” *The Journal of Finance*, 7(1): 77–91.

McCulloch, J. Huston. 2007. “The Kalman foundations of adaptive least squares, with application to US inflation.” Ohio State University Working Paper.

Moench, Emanuel. 2008. “Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach.” *Journal of Econometrics*, 146(1): 26–43.

Moench, Emanuel. 2012. “Term structure surprises: The predictive content of curvature, level, and slope.” *Journal of Applied Econometrics*, 27(4): 574–602.

Moench, Emanuel, Serena Ng, and Simon Potter. 2013. “Dynamic hierarchical factor models.” *Review of Economics and Statistics*, 95(5): 1811–1817.

Neftci, Salih. 2004. *Principles of Financial Engineering.* Academic Press Inc.

Nelson, Charles R., and Andrew F. Siegel. 1987. “Parsimonious modeling of yield curves.” *Journal of Business*, 60(4): 473–489.

Novy-Marx, Robert. 2014. “Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars.” *Journal of Financial Economics*, 112(2): 137–146.

Orphanides, Athanasios, and Min Wei. 2012. “Evolving macroeconomic perceptions and the term structure of interest rates.” *Journal of Economic Dynamics and Control*, 36(2): 239–254.

Pástor, L’uboš, and Robert F Stambaugh. 1999. “Costs of equity capital and model mispricing.” *The Journal of Finance*, 54(1): 67–121.

Pástor, L’uboš, and Robert F Stambaugh. 2000. “Comparing asset pricing models: an investment perspective.” *Journal of Financial Economics*, 56(3): 335–381.

Pettenuzzo, Davide, Allan Timmermann, and Rossen Valkanov. 2014. "Forecasting stock returns under economic constraints." *Journal of Financial Economics*, 114(3): 517–553.

Piazzesi, Monika. 2010. "Affine term structure models." *Handbook of Financial Econometrics*, 1: 691–766.

Piazzesi, Monika, and Martin Schneider. 2007. "Equilibrium yield curves." In *NBER Macroeconomics Annual 2006, Volume 21*. 389–472. MIT Press.

Pooter, Michiel De. 2007. "Examining the Nelson-Siegel class of term structure models." Tinbergen Institute Discussion Paper.

Prado, R., and M. West. 2010. *Time Series: Modeling, Computation, and Inference. Chapman & Hall/CRC Texts in Statistical Science*, Taylor & Francis.

Raftery, Adrian E., Miroslav Kárnỳ, and Pavel Ettler. 2010. "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill." *Technometrics*, 52(1): 52–66.

Rapach, David E., Jack K. Strauss, and Guofu Zhou. 2009. "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy." *Review of Financial Studies*, 23(2): 821–862.

Rossi, Barbara. 2005. "Optimal tests for nested model selection with underlying parameter instability." *Econometric theory*, 21(05): 962–990.

Sangvinatsos, Antonios, and Jessica A. Wachter. 2005. "Does the failure of the expectations hypothesis matter for long-term investors?" *The Journal of Finance*, 60(1): 179–230.

Sargent, Thomas J. 2002. *The Conquest of American Inflation.* Princeton University Press.

Sarno, Lucio, Daniel L. Thornton, and Giorgio Valente. 2007. "The empirical failure of the expectations hypothesis of the term structure of bond yields." *Journal of Financial and Quantitative Analysis*, 42(01): 81–100.

Sarno, Lucio, Paul Schneider, and Christian Wagner. 2014. "The economic value of predicting bond risk premia: Can anything beat the expectations hypothesis." Cass Business School and Centre for Economic Policy Research (CEPR) Working Paper.

Sharpe, William F. 1970. *Portfolio Theory and Capital Markets.* Vol. 217, McGraw-Hill New York.

Siegel, Andrew F., and Charles R. Nelson. 1988. "Long-term behavior of yield curves." *Journal of Financial and Quantitative Analysis*, 23(01): 105–110.

Stock, James H., and Mark W. Watson. 2002. "Forecasting using principal components from a large number of predictors." *Journal of the American Statistical Association*, 97(460): 1167–1179.

Tang, Huarong, and Yihong Xia. 2007. "An international examination of affine term structure models and the expectations hypothesis." *Journal of Financial and Quantitative Analysis*, 42(01): 41–80.

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

Timmermann, Allan G. 1993. "How learning in financial markets generates excess volatility and predictability in stock prices." *The Quarterly Journal of Economics*, 108(4): 1135–1145.

Timmermann, Allan G. 1996. “Excess volatility and predictability of stock prices in autoregressive dividend models with learning.” *The Review of Economic Studies*, 63(4): 523–557.

Ulrich, Maxim. 2013. “Inflation ambiguity and the term structure of us government bonds.” *Journal of Monetary Economics*, 60(2): 295–309.

Uppal, Raman, and Tan Wang. 2003. “Model misspecification and underdiversification.” *The Journal of Finance*, 58(6): 2465–2486.

Van Dijk, Dick, Siem Jan Koopman, Michel Van der Wel, and Jonathan H. Wright. 2014. “Forecasting interest rates with shifting endpoints.” *Journal of Applied Econometrics*, 29(5): 693–712.

Wan, Eric A., and Rudolph Van Der Merwe. 2000. “The unscented Kalman filter for nonlinear estimation.” In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*. 153–158. IEEE.

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

West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag New York, Inc.

Wheelock, David C., and Mark E. Wohar. 2009. “Can the term spread predict output growth and recessions? A survey of the literature.” *Federal Reserve Bank of St. Louis Review*, 91(Part 1): 419–440.

Wright, Jonathan H. 2006. “The yield curve and predicting recessions.” Divisions of Research and Statistics and Monetary Affairs, Federal Reserve Board Working Paper 07.

Wright, Jonathan H. 2011. “Term premia and inflation uncertainty: Empirical evidence from an international panel dataset.” *American Economic Review*, 101(4): 1514–1534.

Appendices

C.1 Bond Pricing in GDTSMs

Under the assumptions in Section 4.2.1, the price of an m-period zero-coupon bond is given by

$$D_t^m = E_t^{\mathbb{Q}}[e^{-\sum_{i=1}^{m-1} r_{t+i}}] = e^{\mathcal{A}_m + \mathcal{B}_m \cdot \mathcal{P}_t}, \tag{50}$$

where $(\mathcal{A}_m, \mathcal{B}_m)$ solve the first-order difference equations

$$\mathcal{A}_{m+1} - \mathcal{A}_m = (K_0^{\mathbb{Q}})' \mathcal{B}_m + \frac{1}{2} \mathcal{B}_m' \Sigma_{\mathcal{P}\mathcal{P}} \mathcal{B}_m - \rho_0, \tag{51}$$

$$\mathcal{B}_{m+1} - \mathcal{B}_m = (K_1^{\mathbb{Q}})' \mathcal{B}_m - \rho_1, \tag{52}$$

subject to the initial conditions $\mathcal{A}_0 = 0$, $\mathcal{B}_0 = 0$. The loadings for the corresponding bond yield are $A_m = -\mathcal{A}_m / m$ and $B_m = -\mathcal{B}_m / m$. See Dai and Singleton (2003) for details.

C.2 Estimation Methods

C.2.1 Bayesian Kalman Filter with Forgetting Factor

We conduct the Kalman filter estimation for the state space model with Equation (4.14) and Equation (4.15):

$$z_t = X_t \beta_t + v_t,$$

$$\beta_{t+1} = \beta_t + u_t,$$

where z_t is an $n \times 1$ vector of variables, $X_t = I_n \otimes \left[z'_{t-1}, ..., z'_{t-p}\right]'$, β_t are VAR coefficients, $v_t \sim N(0, \Sigma_t)$ with Σ_t an $n \times n$ covariance matrix, and $u_t \sim N(0, Q_t)$.

Given that all the data from time 1 to t denoted as D_t, the Bayesian solution to updating about the coefficients β_t takes the form

$$\begin{aligned} p(\beta_t | D_t) &\propto \mathbf{L}(\beta_t; z_t)\, p(\beta_t | D_{t-1}), \\ p(\beta_t | D_{t-1}) &= \int_{\wp} p(\beta_t | D_{t-1}, \beta_{t-1})\, p(\beta_{t-1} | D_{t-1})\, d\beta_{t-1}, \end{aligned}$$

where $\wp$ is the support of β_{t-1}. The solution to this problem can be defined using a Bayesian generalization of the typical Kalman filter recursions. Given an initial condition $\beta_0 \sim N(m_0, \Phi_0)$ we can define (cf. West and Harrison (1997))[34]:

1. Posterior at time $t-1$
$$\beta_{t-1} | D_{t-1} \sim N(m_{t-1}, \Phi_{t-1}),$$

2. Prior at time t
$$\beta_t | D_{t-1} \sim N\left(m_{t|t-1}, \Phi_{t|t-1}\right),$$
where $m_{t|t-1} = m_{t-1}$ and $\Phi_{t|t-1} = \Phi_{t-1} + Q_t$.

3. Posterior at time t
$$\beta_t | D_t \sim N(m_t, \Phi_t), \tag{53}$$
where $m_t = m_{t|t-1} + \Phi_{t|t-1} X'_t (V_t^{-1})' \widetilde{v}_t$ and $\Phi_t = \Phi_{t|t-1} - \Phi_{t|t-1} X'_t (V_t^{-1})' X_t \Phi'_{t|t-1}$, with $\widetilde{v}_t = z_t - X_t m_{t|t-1}$ the prediction error and $V_t = X_t \Phi_{t|t-1} X'_t + \Sigma_t$ its covariance matrix.

Following the discussion above, we need to find estimates for Σ_t and Q_t in the formulas above. We define the time t prior for Σ_t to be

$$\Sigma_t | D_{t-1} \sim iW(S_{t-1}, \delta n_{t-1}), \tag{54}$$

[34]For a parameter θ we use the notation $\theta_{t|s}$ to denote the value of parameter θ_t given data up to time s (i.e. D_s) for $s > t$ or $s < t$. For the special case where $s = t$, I use the notation $\theta_{t|t} = \theta_t$

while the posterior takes the form

$$\Sigma_t|D_t \sim iW\left(S_t, n_t\right),$$

where $n_t = \delta n_{t-1} + 1$ and $S_t = \delta S_{t-1} + n_t^{-1}\left(S_{t-1}^{0.5} V_{t-1}^{-0.5} \widetilde{v}_{t|t-1} \widetilde{v}'_{t|t-1} V_{t-1}^{-0.5} S_{t-1}^{0.5}\right)$. In this formulation, v_t is replaced with the one-step-ahead prediction error $\widetilde{v}_{t|t-1} = z_t - m_{t|t-1}X_t$. The estimate for Σ_t is approximately equivalent numerically to the Exponentially Weighted Moving Average (EWMA) filter $\hat{\Sigma}_t = \delta\hat{\Sigma}_{t-1} + (1-\delta)\, v_t v_t'$. The parameter δ is the decay factor, where for $0 < \delta < 1$. In fact, Koop and Korobilis (2013) apply such a scheme directly to the covariance matrix Σ_t, which results in a point estimate. In this case by applying variance discounting methods to the scale matrix S_t, we are able to approximate the full posterior distribution of Σ_t.

Regarding Q_t, we use the forgetting factor approach in Koop and Korobilis (2013); see also West and Harrison (1997) for a similar discounting approach. In this case Q_t is set to be proportionate to the filtered covariance $\Phi_{t-1} = cov\left(\beta_{t-1}|D_{t-1}\right)$ and takes the following form

$$Q_t = \left(\lambda^{-1} - 1\right)\Phi_{t-1}, \tag{55}$$

for a given forgetting factor λ. Note that λ is mathematically equivalent to the quantity $1-\gamma$ in the *constant gain least squares* (CGLS) algorithm, see Appendix C.2.2 and McCulloch (2007). Therefore, the forgetting factor λ and the gain parameter γ are two sides to the same coin. As λ becomes larger, the γ becomes smaller, so the model would adjust more slowly if a structural break happens.

An alternative brief interpretation of forgetting factors is that they control how much 'recent past' information will be used. With the exponential decay for the forgetting factors, if it takes a value of 0.99, the information 24 periods ago (two years for monthly data) receives around 80% as much weight as the information of last period. If forgetting factor takes 0.95, then forecast performance 24 periods ago receives only about 30% as much weight. The similar implication holds for the decay factor.

C.2.2 The Link between the Kalman Filter and Adaptive Least Squares

From the Kalman filter described in last section, we have the following formulas

$$m_t = m_{t|t-1} + \Phi_{t|t-1}X_t'(V_t^{-1})'\widetilde{v}_t, \tag{56}$$

$$\Phi_t = \Phi_{t|t-1} - \Phi_{t|t-1}X_t'(V_t^{-1})'X_t\Phi_{t|t-1}', \tag{57}$$

$$V_t = X_t\Phi_{t|t-1}X_t' + \Sigma_t, \tag{58}$$

where $\widetilde{v}_t = z_t - X_t m_{t|t-1}$ is the prediction error. Post-multiply (57) by X_t' and combine with (58) we obtain

$$\begin{aligned}\Phi_t X_t' &= \Phi_{t|t-1}\big(X_t' - X_t'(V_t^{-1})' X_t \Phi_{t|t-1}' X_t'\big)\\ &= \Phi_{t|t-1}\big(X_t' - X_t'(V_t^{-1})'(V_t - \Sigma_t)\big)\\ &= \Phi_{t|t-1} X_t'(V_t^{-1})'\Sigma_t. \end{aligned} \tag{59}$$

We can get the expressions of ALS by post-multiplying (59) by Σ_t^{-1} and substituting it back to (56) and (57), respectively.

For (56) we have

$$m_t = m_{t|t-1} + R_t^{-1} X_t' \Sigma_t^{-1} \widetilde{v}_t, \tag{60}$$

where we set $R_t = \Phi_t^{-1}$. So we obtain the evolution of the drift in ALS.

We continue the previous substitution in (57) with $\Phi_{t|t-1} = \Phi_{t-1} + Q_t$ in hand, which gives

$$\Phi_t = \Phi_{t-1} + Q_t - \Phi_t X_t' \Sigma_t^{-1} X_t \Phi_{t-1} - \Phi_t X_t' \Sigma_t^{-1} X_t Q_t. \tag{61}$$

Setting $R_t = \Phi_t^{-1}$, we can get the final equation after some manipulation

$$R_t = (I + Q_t \Phi_t^{-1})^{-1} R_{t-1} + X_t' \Sigma_t^{-1} X_t. \tag{62}$$

If we set $Q_t = \frac{\gamma}{1-\gamma}\Phi_{t-1}$, then we have the *constant gain least squares* (CGLS) algorithm.

C.2.3 Brief Introduction of the Unscented Kalman Filter

Consider the following nonlinear discrete-time stochastic system represented by:

$$z_t = f(\beta_t) + v_t, \tag{63}$$

$$\beta_{t+1} = \beta_t + u_t, \tag{64}$$

where z_t is an $n \times 1$ vector of variables, β_t are coefficients that govern the pricing equation $f(\cdot)$, $v_t \sim N(0, \Sigma_t)$ with Σ_t an $n \times n$ covariance matrix, and $u_t \sim N(0, Q_t)$.

As we mentioned before, the solution for this system follows a recursive rule given by

$$\beta_t | D_t \sim N(m_t, \Phi_t), \tag{65}$$

where D_t is the information set at time t. Similar to the Kalman filter, the unscented Kalman filter has the same recursive estimation process as in Appendix C.2.1, except the update equations (56) and

(57) are replaced by:

$$m_t = m_{t|t-1} + \mathcal{K}_t \widetilde{v}_t, \tag{66}$$

$$\Phi_t = \Phi_{t|t-1} - \mathcal{K}_t P_{z_t} \mathcal{K}_t', \tag{67}$$

where $\mathcal{K}_t$ is the Kalman gain of the filter and P_{z_t} is the prior variance of z_t. The above updating equations are intuitively similar to the ones in the Kalman filter, except we use different formulas to obtain the Kalman gain and the prior variance. To be more specific, these quantities can be calculated by simulating sigma points around the mean of state variables, see Wan and Van Der Merwe (2000) and Appendix C.4 for details.

C.2.4 Probabilities for Dynamic Model Selection

To obtain the desire probabilities for dynamic selection or averaging, we need updating at each point in time. In papers such as Raftery, Kárnỳ and Ettler (2010) or Koop and Korobilis (2012) the models are TVP regressions with different sets of explanatory variables. The analogous result of the model prediction equation, when doing Bayesian model averaging, is

$$p(\beta_{t-1}|D_{t-1}) = \sum_{k=1}^{K} p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1}) \mathbf{Pr}(L_{t-1} = k|D_{t-1}), \tag{68}$$

where $L_{t-1} = k$ means the k_{th} model[35] is selected and $p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1})$ is given by the Kalman filter (Equation 53). To simplify notation, let $\pi_{t|s,l} = \mathbf{Pr}(L_t = l|D_s)$.

Raftery, Kárnỳ and Ettler (2010) used an empirically sensible simplification that

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^{\alpha}}{\sum_{l=1}^{K} \pi_{t-1|t-1,l}^{\alpha}}, \tag{69}$$

where $0 < \alpha \leq 1$. A forgetting factor is also employed here, of which the meaning is discussed in the last section.[36] The huge advantage of using the forgetting factor α is that it does not require an MCMC algorithm to draw transitions between models or a simulation algorithm over model space.

The model updating equation is

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} p_k(z_t|D_{t-1})}{\sum_{l=1}^{K} \pi_{t|t-1,l} p_l(z_t|D_{t-1})}, \tag{70}$$

where $p_k(z_t|D_{t-1})$ is the predictive likelihood. When proceeding with dynamic model selection (DMS), the model with the highest probability is the best model we would like to select. Alternatively,

[35] For example, it can be the k_{th} model in a pool of possible models with different restrictions or gain parameter γ.

[36] In this paper, we set $\alpha = 1$ to put equal weights to previous information.

we can conduct dynamic model averaging (DMA), which average the predictions of all models with respective probabilities.

C.3 Proof of Propositions

C.3.1 Heuristics of Proposition 1

Following Garlappi, Uppal and Wang (2007), we start with the inner minimization

$$\min_{\mu} w^{\intercal}\mu - \frac{\gamma}{2} w^{\intercal}\Sigma w, \tag{71}$$

subject to

$$(\hat{\mu}-\mu)^{\intercal}\Sigma^{-1}(\hat{\mu}-\mu) \leq \varepsilon. \tag{72}$$

The Lagrangian is given by

$$\mathcal{L}(\mu,\lambda^{\mathcal{L}}) = w^{\intercal}\mu - \frac{\gamma}{2} w^{\intercal}\Sigma w - \lambda^{\mathcal{L}}[\varepsilon - (\hat{\mu}-\mu)^{\intercal}\Sigma^{-1}(\hat{\mu}-\mu)]. \tag{73}$$

μ^* is a solution of the constrained problem above if and only if there exists a scalar $\lambda^{\mathcal{L}*} \geq 0$, such that $(\mu^*, \lambda^{\mathcal{L}*})$ is a solution of the following unconstrained problem

$$\min_{\mu} \max_{\lambda^{\mathcal{L}}} \mathcal{L}(\mu,\lambda^{\mathcal{L}}). \tag{74}$$

From the first order conditions with respect to μ in Equation (73), we have

$$\mu^* = \hat{\mu} - \frac{1}{2\lambda^{\mathcal{L}}}\Sigma w. \tag{75}$$

Substituting this in the Lagrangian (73) we obtain

$$\mathcal{L}(\mu^*,\lambda^{\mathcal{L}}) = w^{\intercal}\hat{\mu} - (\frac{1}{4\lambda^{\mathcal{L}}} + \frac{\gamma}{2}) w^{\intercal}\Sigma w - \lambda^{\mathcal{L}}\varepsilon. \tag{76}$$

Therefore, the original max-min problem with constraints is equivalent to the maximization problem below

$$\max_{w,\lambda^{\mathcal{L}}} w^{\intercal}\hat{\mu} - (\frac{1}{4\lambda^{\mathcal{L}}} + \frac{\gamma}{2}) w^{\intercal}\Sigma w - \lambda^{\mathcal{L}}\varepsilon. \tag{77}$$

Solving for $\lambda^{\mathcal{L}}$, we get $\lambda^{\mathcal{L}} = \frac{1}{2}\sqrt{\frac{w^{\intercal}\Sigma w}{\varepsilon}} > 0$. Then we can rewrite the maximization problem as

$$\max_{w} w^{\intercal}\hat{\mu} - \frac{\gamma}{2} w^{\intercal}\Sigma w (1 + \frac{2\sqrt{\varepsilon}}{\gamma\sqrt{w^{\intercal}\Sigma w}}). \tag{78}$$

It is easy to show, the first-order condition with respect to w gives

$$w = (\frac{\sigma_p}{\gamma\sigma_p + \sqrt{\varepsilon}})\Sigma^{-1}\hat{\mu}. \tag{79}$$

With Equation (79), we can post-multiply $w^\intercal$ by Σw and obtain

$$\sigma_p^2 = \big(\frac{\sigma_p}{\gamma\sigma_p + \sqrt{\varepsilon}}\big)^2 \hat{\mu}^\intercal \Sigma^{-1} \hat{\mu}, \tag{80}$$

where $\sigma_p = \sqrt{w^\intercal \Sigma w}$.

After some manipulation, the optimal portfolio weight w^* is given by the positive real solution σ_p^* of the following polynomial

$$\gamma^2 \sigma_p^2 + 2\sqrt{\varepsilon}\gamma\sigma_p + \varepsilon - \hat{\mu}^\intercal \Sigma^{-1} \hat{\mu} = 0. \tag{81}$$

If $\hat{\mu}^\intercal \Sigma^{-1} \hat{\mu}$ is sufficiently large, we have a unique positive real solution σ_p^*. Otherwise, we have a non-negative solution σ_p^*, i.e. $w^* = \mathbf{0}$. Therefore, using Equation (79) we have Equation (4.29) in **Proposition 1**.

C.3.2 Heuristics of Proposition 2

To solve the max-min problem

$$\max_{w,\hat{\mu}} \min_{\mu} w^\intercal \mu - \frac{\gamma}{2} w^\intercal \Sigma w, \tag{82}$$

subject to

$$(\hat{\mu} - \mu)^\intercal \Sigma^{-1} (\hat{\mu} - \mu) \leq \varepsilon, \tag{83}$$

$$\hat{\mu} \in \{\hat{\mu}_k : k = 1, ..., K\}, \tag{84}$$

we follow the same procedures as in Appendix C.3.1. The difference lies in the outer maximization:

$$\max_{w,\hat{\mu}} w^\intercal \hat{\mu} - \frac{\gamma}{2} w^\intercal \Sigma w \big(1 + \frac{2\sqrt{\varepsilon}}{\gamma\sqrt{w^\intercal \Sigma w}}\big), \tag{85}$$

where we need to consider first-order conditions with respect to $\hat{\mu}$ as well as w. We have the same formula (79) for w. However, in addition to w, we need to search $\hat{\mu}$ over a set of possible models at each point in time, and use the optimal forecasts $\hat{\mu}^*$ that give the largest value in the above maximization problem.

C.4 Technical Details of the Unscented Kalman Filter

C.4.1 Unscented Transformation

The UKF is based on the *unscented transformation* (UT) in order to form a Gaussian approximation to the target distribution. The advantage of UT over the Taylor series based approximation in other nonlinear filters (for example, the extended Kalman filter) is that Jacobian and Hessian matrices are not need, so the estimation procedure is more convenient in a system where closed-form expressions are not available.

The follows show the procedure of unscented transformation:

1. We simulate a set of $2n+1$ sigma points $\mathcal{X}$ of the state variables x, where n is the dimension of the state, from the mean m and covariance matrix Φ:

$$\begin{aligned} \mathcal{X}^{(0)} &= m, \\ \mathcal{X}^{(i)} &= m + \sqrt{(n+\lambda)\Phi}, \quad i = 1, ..., n, \\ \mathcal{X}^{(i)} &= m - \sqrt{(n+\lambda)\Phi}, \quad i = n+1, ..., 2n, \end{aligned} \tag{86}$$

   with the associated weights W_m of the state variables x and W_c of the observations z:

$$\begin{aligned} W_m^{(0)} &= \frac{\lambda}{n+\lambda}, \\ W_c^{(0)} &= \frac{\lambda}{n+\lambda} + (1-\alpha^2+\beta), \\ W_m^{(i)} = W_c^{(i)} &= \frac{1}{2(n+\lambda)}, \quad i = 1, ..., 2n, \end{aligned} \tag{87}$$

   where $\lambda = \alpha^2(n+\kappa) - n$ is a scaling parameter. α and κ determine the spread of the sigma points around the state, and β is used to incorporate prior knowledge of the distribution of the state.[37]

2. The sigma points are propagated though non-linearity as

$$\mathcal{Z}^{(i)} = f(\mathcal{X}^{(i)}), \quad i = 0, ..., 2n. \tag{88}$$

[37] As suggested by Wan and Van Der Merwe (2000), normal values are $\alpha = 10^{-3}$, $\kappa = 0$ and $\beta = 2$. If the true distribution of x is Gaussian, $\beta = 2$ is optimal. Note that the simple approximation approach taken with the UT are accurate to the third order for all nonlinearities with Gaussian innovations, which has an advantage over Monte-Carlo methods which require (orders of magnitude) more sample points to provide an accurate distribution of the state.

3. We can compute the mean and covariance estimates for z:

$$\bar{z} \approx \sum_{i=0}^{2n} W_m^{(i)} \mathcal{Z}^{(i)},$$
$$P_z \approx \sum_{i=0}^{2n} W_c^{(i)} (\mathcal{Z}^{(i)} - \bar{z})(\mathcal{Z}^{(i)} - \bar{z})^{\mathsf{T}}. \tag{89}$$

4. Estimation of the cross-covariance between z and x is given by

$$P_{x,z} \approx \sum_{i=0}^{2n} W_c^{(i)} (\mathcal{X}^{(i)} - m)(\mathcal{Z}^{(i)} - \bar{z})^{\mathsf{T}}. \tag{90}$$

C.4.2 Estimation Procedure using UKF

Based on the discussion of UT above, we describe the following *prediction* and *update* steps of the UKF.

- *Prediction*: Compute the predicted state mean $m_{t|t-1}$ and covariance $\Phi_{t|t-1}$, the predicted observation mean $\hat{Z}_t$ and covariance P_{z_t}, and the cross-variance of the state and measurement P_{x_t,z_t}:

$$\begin{aligned}
m_{t|t-1} &= m_{t-1|t-1}, \\
\Phi_{t|t-1} &= \Phi_{t-1|t-1} + Q_t, \\
\hat{z}_t &= \sum_{i=0}^{2n} W_m^{(i)} \mathcal{Z}_{t|t-1}^{(i)}, \\
P_{z_t} &= \sum_{i=0}^{2n} W_c^{(i)} (\mathcal{Z}_{t|t-1}^{(i)} - \hat{z}_t)(\mathcal{Z}_{t|t-1}^{(i)} - \hat{z}_t)^{\mathsf{T}} + \Sigma_t, \\
P_{x_t,z_t} &= \sum_{i=0}^{2n} W_c^{(i)} (\mathcal{X}_{t|t-1}^{(i)} - m_{t|t-1})(\mathcal{Z}_{t|t-1}^{(i)} - \hat{z}_t)^{\mathsf{T}}.
\end{aligned} \tag{91}$$

- *Update*: Compute the filter gain $\mathcal{K}_t$ and the updated state mean $m_{t|t}$ and covariance $\Phi_{t|t}$ in order to get Equations (66) and (67):

$$\begin{aligned}
\mathcal{K}_t &= P_{x_t,z_t} P_{z_t}^{-1}, \\
m_{t|t} &= m_{t|t-1} + \mathcal{K}_t \widetilde{v}_t, \\
\Phi_{t|t} &= \Phi_{t|t-1} - \mathcal{K}_t P_{x_t} \mathcal{K}_t'.
\end{aligned} \tag{92}$$

Following Koop and Korobilis (2012) and Koop and Korobilis (2013), we specify

$$Q_t = \left(\lambda_f^{-1} - 1\right) \Phi_{t-1}, \tag{93}$$

where λ_f is the ‘forgetting factor’. We have an intuitive interpretation for the forgetting factor:

the smaller the λ_f, the more weights UKF puts on the new information, and hence the system is more sensitive to structural changes.[38] To fix the idea, we set the value to 0.99 to ensure the stability of the system.

C.4.3 Detailed Specification of the ATSM

We adopt a specific parametric form of the class of Affine Term Structure Models (ATSMs) with arbitrage-free restrictions under the Duffie and Kan (1996) framework, which is similar to the majority of current related literature, see for example, Duffee (2002), Dai and Singleton (2003), Joslin, Singleton and Zhu (2011) and Joslin, Priebsch and Singleton (2014).[39] In this setup, the measurement equation in the nonlinear system is governed by parameter set $(\Sigma_{\mathcal{PP}}, \lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$.[40] Joslin, Singleton and Zhu (2011) prove that every canonical GDTSM is observationally equivalent to the canonical GDTSM:

$$X_t = J(\lambda^{\mathbb{Q}}) + \sqrt{\Sigma_X}\varepsilon_t^{\mathbb{Q}}, \tag{94}$$

$$r_t = r_\infty^{\mathbb{Q}} + \mathbf{1} \cdot X_t, \tag{95}$$

where X_t are normalized risk factors, $r_\infty^{\mathbb{Q}}$ denotes the unconditional mean of r_t under $\mathbb{Q}$, and $J(\lambda^{\mathbb{Q}})$ is a real Jordan form matrix associated with eigenvalues $\lambda^{\mathbb{Q}}$. We can conveniently apply invariant transformation of X_t and then replace the risk factors with preferred portfolio combinations, see Dai and Singleton (2000) and Joslin, Singleton and Zhu (2011) for details.

Solving for the bond prices of m-period zero-coupon bond D_t^m using the recursion given by

$$D_t^m = E_t^{\mathbb{Q}}[e^{-\sum_{i=1}^{m-1} r_{t+i}}], \tag{96}$$

we can obtain the following pricing equation for m-period bond yields as the measurement equation:

$$y_t^m = A_X^m(\Sigma_{t,\mathcal{PP}}, \lambda_t^{\mathbb{Q}}, r_{t,\infty}^{\mathbb{Q}}) + B_X^m(\lambda_t^{\mathbb{Q}})X_t, \tag{97}$$

where

$$\begin{aligned} A_X^{m+1} - A_X^m &= \frac{1}{2} B_X^{m\prime}\, \Sigma_X B_X^m - r_{t,\infty}^{\mathbb{Q}}, \\ B_X^{m+1} - B_X^m &= J(\lambda_t^{\mathbb{Q}})' B_X^m. \end{aligned} \tag{98}$$

[38]To see this, use Taylor series expansion around the observation mean. See Koop and Korobilis (2012) and Koop and Korobilis (2013) for detailed discussion about the ‘forgetting factor’.

[39]Joslin, Singleton and Zhu (2011) denote their proposed model as a canonical Gaussian dynamic term structure model (GDTSM). We use ATSM and GDTSM interchangeably.

[40]Following the notations in Joslin, Singleton and Zhu (2011) and Joslin, Priebsch and Singleton (2014), $\lambda^{\mathbb{Q}}$ denotes the N-vector of ordered nonzero eigenvalues of $K_{\mathcal{PP}}^{\mathbb{Q}}$ and $r_\infty^{\mathbb{Q}}$ denotes the long-run mean of r_t under $\mathbb{Q}$.

CHAPTER 5

Conclusion

5.1 Summary and Policy Implications

The topic of the term structure of interest rates is of importance to investors and to policymakers who wish to extract macroeconomic expectations from the term structure and take action to affect the interest rates of different maturities. Researchers are interested in term structure models for many reasons. First, term structure provides information about the expectations of financial market participants, which are of great interest to forecasters and policymakers. The decisions of economic agents are influenced by their expectations, and these expectations, in turn, help determine what actually happens in the future. This *self-fulfilling prophecy* is critical for forecasting future paths of monetary policies and evaluating the effects, as well as for predicting asset returns and portfolio allocations of investors and for their strategies for hedging interest rate risk. Allowing for a large dataset could be helpful in terms of forecasting, but a large model is not always superior as we have shown in Chapter 2. We need be careful with the implementation in order to properly assimilate useful information.

Second, economists are interested in term structure theories because they have implications for how monetary policy should respond to changes in long-term interest rates. Economic theory suggests that monetary policy may have a direct effect on short-term interest rates, but may not directly affect long-term rates. A number of important economic decisions are related to long-term rates, such as firms' decisions about investment, and individuals' decisions about the purchase of homes and other durable goods. Gürkaynak and Wright (2012) suggest that if the fall in long rates is due to the fall in risk premia, policymakers ought to *lean against the wind* by tightening the stance of monetary policy to offset the additional stimulus to aggregate demand. However, the determination of endogenous bond yields is still not very clear, and further research is needed to understand the transmission mechanism of monetary policies.

Third, the federal funds rate hit the zero lower bound during the financial crisis and as a result policymakers may wish to provide additional stimulus to the economy. Under the simple expectations hypothesis with constant term premia, policymakers can influence the market participants' expectations about future monetary policy by committing to keeping the federal funds rate at zero for an extended period. However, this simple expectations theory has been rejected many times in careful econometric studies, and it is imperative to take account of time-varying premia. If long-term interest rates are also buffeted by risk premia, then measures to alter those risk premia at the zero bound (for example, large-scale asset purchases) may be effective. Therefore, more sophisticated term structure models beyond the expectations hypothesis are by all means desirable. Moreover, when confronting the expectations hypothesis, it is interesting to understand how economic agents can benefit from more sophisticated term structure models when there is ambiguity about the correct model.

This thesis provides an analysis of the term structure of interest rates related to the above questions, with special emphasis on the intersection of macroeconomics and finance with the applications of Bayesian econometrics. In three distinct but logically interconnected Bayesian settings I demonstrate i) how to properly incorporate a large set of macro-finance information to increase forecasting performance, while considering time-varying coefficients, stochastic volatility and dynamic model averaging; ii) the underlying sources influencing the global comovement of bond yields and the underlying economic mechanism; and iii) what drives interest rate predictability, and how an economic agent makes portfolio choices in the presence of model uncertainty, in order to have significant gains over the expectations hypothesis benchmark.

In Chapter 2, we extend the dynamic Nelson-Siegel approach of yield curve modeling by employing Dynamic Model Averaging (NS-DMA), in order to characterize the nonlinear dynamics of yield factors, as Duffee (2002) suggests nonlinearity can potentially improve yield forecasts. The framework we propose generalizes some frontier econometric techniques, and is augmented with many (unspanned) macro-finance factors, as in Dewachter and Iania (2012). The NS-DMA method significantly improves the predictive accuracy and successfully identifies the dynamics of term premia, on the grounds that it seems to have appropriately incorporated the information in the macro-economy. Our approach allows for potential structural breaks and model uncertainty, and the out-of-sample predictability of our real-time forecast model is statistically significant when compared with benchmark models. We specifically discuss some informative responses of bond yields to monetary policy implementations in different periods, such as the Great Moderation and the financial crisis. The term premia is generally countercyclical but a distinct 'flight-to-quality' demand in the financial crisis is revealed.

In Chapter 3, we propose a new 'Fundamentals-Augmented Hierarchical Factor Model' to jointly identify global and national Level and Slope factors augmented with global fundamentals: inflation, real activity, changes in policy rate and financial conditions. Co-movement accounts for, on average, two thirds of the variability in global bond yields. Our method is robust to the macro spanning condition and is able to recover significant explanatory power of global inflation shocks for global yield co-movement, through a policy channel and a risk compensation channel. Shocks to non-fundamentals are persistent and account for the majority of global term structure movement. Moreover, we find that the non-fundamental movements can be satisfactorily explained by measures of sentiment and economic uncertainty. Country-specific components contribute to the majority of remaining variance, of which one fourth is due to spillovers. Therefore, the latent information used to price bond yields has economic appeal and a structural model considering these salient features is promising in explaining the underlying economic mechanism.

Chapter 4 studies the problem of a Bayesian representative agent who learns about the information in the bond market over time, with the consideration of parameter uncertainty and model uncertainty. In addition to adaptive learning about parameters considered in Giacoletti, Laursen and Singleton (2014), our proposed framework provides flexibility in specifying different learning speeds and model restrictions. We can select the optimal specification according to predictive performance over time, and hence reduce the risk of data snooping problem. This method reveals the agent's expectations in real time by using conditional information. We find that apart from observational variance, parameter instability is the dominant driving force of predictive uncertainty, when compared with uncertainty in learning speed or model restrictions. It suggests that a successful term structure model should at least consider time-varying parameters when making conditional forecasts.

Lastly, in Chapter 4 we study the problem of asset allocation for an investor with ambiguity aversion building upon Garlappi, Uppal and Wang (2007). After learning the parameters, our ambiguity-averse investor forms optimal portfolios by maximizing mean-variance expected utility. The ensemble of all salient features offered by our framework is essential in producing significant and consistent economic value over the Expectations Hypothesis benchmark. Intuitively, ambiguity aversion with model uncertainty ensures that we refine the search for portfolio weights in a reliable region, which in turn not only boosts but also stabilizes the gains. We are convinced that ambiguity aversion is key to salvaging the models with significant predictability but little economic value in the previous literature, and the *economic value puzzle* in bond returns can be resolved following this direction.

5.2 Further Research

Applying sophisticated Bayesian econometric techniques is promising in exploring the frontier of term structure modeling. However, the curse of dimensionality and the tractability of a complicated system are important problems when pushing the boundaries of this research area. To advance the topics discussed in Chapter 2, future research could employ a one-step approach to provide forecasts with higher accuracy, in which case a trade-off should be made between predictive accuracy and estimation efficiency. Discussing the real part of the term structure is meaningful and desirable, but it requires a more complicated specification and the estimation would be challenging.

Building upon Chapter 3, it is interesting to specifically evaluate whether the contagion across different countries is related to fundamental or nonfundamental drivers, and a much larger system is needed to fulfill this goal. The method in this chapter does not explicitly model potential time-varying nonlinear dynamics of yield factors, such as regime shifts. Allowing for nonlinearity can be promising in unfolding more informative dynamics of fundamental and non-fundamental fluctuations.

For Chapter 4, by allowing for more general model specifications, such as incorporating more information from macro-finance predictor variables or economic constraints in Pettenuzzo, Timmermann and Valkanov (2014), we should be able to further improve model performance and provide meaningful economic rationales. It would also be interesting to develop hybrid models with both spanned and unspanned macroeconomic risks and explore the prediction uncertainty from different predictors, as suggested by Bauer and Rudebusch (2015). Lastly, our results suggest that the zero lower bound problem could hinder the performance of our portfolio strategy. The above potential extensions are not only economically meaningful, but also raise challenging econometric obstacles which we aim to tackle in the future.